DEMONSTRATION REPORT

Dipole Models for UXO Discrimination at Live Sites – Pole Mountain

ESTCP Project MR-201159

JUNE 2012

Leonard Pasion
Sky Research, Inc.

ESTCP

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Dipole Models for UXO Discrimination at Live Sites - Pole Mountain	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Sky Research, Inc.,445 Dead Indian Memorial Road,Ashland,OR,97520**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**59**	

Executive Summary

The demonstration described in this report was conducted at the Pole Mountain, Wyoming, under project ESTCP MR-201159 "Dipole Models for UXO discrimination at Live Sites." It was performed under the umbrella of the ESTCP Discrimination Study Program. The objective of the MR-201159 project is to demonstrate the application of feature extraction and statistical classification to the problem of UXO classification. At the Pole Mountain site, the objective was to discriminate targets of interest (TOI) (including 37 mm, 57 mm, 60 mm, 75 mm targets, a stokes mortar and a small industry standard object (ISO)) from non-hazardous shrapnel, range and cultural debris. In this report, we describe the performance of classification techniques that utilized data from a Geonics EM61 cart deployed in full coverage and MetalMapper data acquired in a cued interrogation mode.

The classification techniques applied to the Pole Mountain data use features extracted from a dipole model fit to the observed data. Features used were the dipole polarizability tensor for the MetalMapper data and measures of the polarizability magnitude and decay rate for the Geonics EM61 data. From the extracted feature vectors the following prioritized dig-lists were created: (i) EM61 cart data ranked by polarizability time-decay and data amplitude; (ii) MetalMapper statistical classification using a Support Vector Machine applied to polarizabilities; and (iii) MetalMapper classification using a library matching method. All model fits and discrimination analysis were performed using the Sky classification software suite (UXOLab) that was jointly developed by UBC-GIF and Sky Research.

The blind-test for the MetalMapper was split into two parts: Year 1 and Year 2. The Year 1 data set consisted of anomalies in the Southern portion of the site and the Year 2 data set consisted of anomalies in the Northern portion of the site. Year 1 targets were analyzed and scored first. The results from the Year 1 analysis were provided to data processors, and were used when processing the Year 2 anomalies.

The EM-61 cart data were inverted for dipole parameters, from which polarizability time decay (based on the ratio of the total polarizability at the first and fourth time gates) was derived. As a proxy for size, we used the maximum amplitude of the first time channel of the data anomaly. To classify the EM-61 data, we trained a probabilistic neural network (PNN) classifier in the two-dimensional data amplitude/decay feature space. The Year 1 test contained 984 anomalies (46 TOI and 938 non-TOI). At the operating point 650 excavations were required and all TOI were recovered, along with 604 of 938 non-TOI (64.4% of the clutter). The Year 2 test contained 1384 anomalies (114 TOI and 1270 non-TOI). At the operating point 650 excavations were required and all TOI were recovered, along with 945 of 1270 non-TOI (74.4% of the clutter). The false alarm rates (FAR) for Year 1 and Year 2 were, 64.0% and 68.0% respectively. The mediocre performance of the EM-61 can be attributed to the relative size of the smallest target of interest (the small ISO) and scrap, and the data's inability to accurate constrain their size and shape. The operating point (or stop-dig point) for both Year 1 and Year 2 were chosen such that no TOI were left in the ground.

For both the Year 1 and Year 2 MetalMapper data sets, an SVM based statistical classifier dig-list and a library based classifier dig-list were submitted. The SVM statistical classifier and library based classifier produced similar results. For the Year one test, the Library method resulted in the excavation of all TOI and 67 of 938 non-TOI (7.14% of clutter), with an FAR of 2.77%. Application of the SVM classifier

resulted in the excavation of all TOI and 114 of 938 non-TOI (12.2 % of clutter) and a FAR of 3.62%. The Year 2 test again produced similar results between the SVM statistical classifier and library based classifier. The Library method resulted in the excavation of all TOI and 102 of 1270 non-UXO (8.03% of clutter) and a FAR of 2.20%. Application of the SVM classifier resulted in the excavation of all TOI and 291 of 1270 non-TOI (22.91% of clutter) with a FAR of 1.73%. The additional digs for the SVM classifier can be attributed to the more conservative, automated method for determining the stop-dig point. Ordnance type was predicted using the Library matching technique. The correct ordnance type was predicted in 156 of 160 cases (97.5% success rate). Each of the four incorrect caliber estimates were from 57 mm mortars incorrectly classified as 60 mm mortars.

In this report, we also present results achieved through technology transfer with Shaw Environmental. Two Shaw personnel visited the Vancouver Sky Research office for a one week training session. The training consisted of an overview of the different aspects of UXO classification: the dipole model, data inversion, data and inversion quality control checks, training data selection, classification methods, and the use of classification software. With guidance and assistance from Sky Research analysts, they processed the combined Years 1 and 2 dataset and developed a dig list using Sky Research software for data inversion, QC, selection of training data and creation of the dig list. Of the 2370 anomalies, a total of 236 anomalies were chosen for investigation. At the operating point all TOI and 97 of the 2208 non-TOI (4.40% of clutter) were excavated. The FAR was 2.40%.

There are three conclusions from applying dipole-based classification techniques to the Pole Mountain demonstration. Firstly, the production quality EM61 data produced mediocre results due to the relative sizes of the scrap and the smallest UXO, and the inability to accurate constrain their size and shape. Secondly, the MetalMapper sensor deployed in a cued-interrogation mode resulted in a significant reduction in the number of digs required to excavate all UXO, and an ability to accurately distinguish different UXO types from one another. Finally, excellent classification results using MetalMapper data from Pole Mountain were achieved by visiting Shaw Environmental geophysicists after a weeklong training session and use of the Sky classification software suite.

Table of Contents

List of Tables

List of Figures

Acronyms

API	Application Programming Interface
EM	Electromagnetic
EMI	Electromagnetic induction
ESTCP	Environmental Security Technology Certification Program
FOM	Figures of Merit
GPS	Global Positioning Systems
IDA	Institute for Defense Analysis
IMU	Inertial Measurement Unit
PNN	Probabilistic Neural Network
QC	Quality Control
SLO	San Luis Obispo
SNR	Signal to noise ratio
SVM	Support vector machine
TEM	Time-domain electromagnetic
TEMTADS	Time Domain Electromagnetic Towed Array Detection System
TOI	Targets of Interest
UBC-GIF	University of British Columbia – Geophysical Inversion Facility
USACE-ERDC	United States Army Corps of Engineers-Engineering Research and Development Center
UXO	Unexploded ordnance

1. INTRODUCTION

1.1 BACKGROUND

Significant progress has been made in discrimination technology. To date, testing of these approaches has been primarily limited to test sites with only limited application at live sites. Acceptance of discrimination technologies requires demonstration of system capabilities at real UXO sites under real world conditions. Any attempt to declare detected anomalies to be harmless and requiring no further investigation will require demonstration to regulators of not only individual technologies, but an entire decision making process.

To demonstrate the viability of advanced detection and discrimination technologies, ESTCP has now conducted four UXO classification studies. The results of the first demonstration, at the former Camp Sibert, Alabama were very encouraging. Although conditions were favorable at this site, including a single target-of-interest (4.2-in mortar) and benign topography and geology, all of the demonstrated classification approaches were able to correctly identify a sizable fraction of the anomalies as arising from non-hazardous items that could be safely left in the ground. Of particular note, the contractor EM-61-MK2 cart survey with analysis using commercially available methods correctly identified more than half the targets as non-hazardous.

To build upon the success of this first study, ESTCP expanded the program to include a second study at a site with more challenging topography and a wider mix of targets-of interest. A range at the former Camp San Luis Obispo (SLO), California, was selected for this demonstration. We again found that, with appropriate use of classification metrics applied to production quality EM-61 data, it was possible to significantly reduce the number of clutter items excavated without missing any targets of interest (TOI). Furthermore, the next generation of EM sensors, when deployed in a cued-interrogation mode, produced significant additional reductions in the number of clutter items excavated. These sensors could also usually distinguish between different UXO types. A third ESTCP demonstration study was conducted in 2010 at Camp Butner, North Carolina. The site had very little topographic relief but required classification between small targets of interest (37mm projectiles and M48 fuses) and metallic debris of similar size. Targets were also distributed with a higher density than previously encountered. While production data sets performed poorly at Camp Butner, excellent discrimination performance was achieved with cued data sets. In particular, the TEMTADS array was able to identify all UXO with approximately 5% of non-UXO dug.

1.2 OBJECTIVE OF THE DEMONSTRATION

There were two objectives in this demonstration. The first objective of this demonstration were to perform data modeling, classification, and classification using electromagnetic (EM) data collected at Pole Mountain, Wyoming. We processed two data sets:

1. EM-61 cart data;
2. MetalMapper EMI sensor cued interrogation data;

For the EM61 cart data, we submitted a single diglist based on polarizability time decay and data amplitude. Anomalies were prioritized using a probabilistic neural network. For the MetalMapper data set, polarizabilities were derived and used for classification. A support vector machine and rule-based classifier were used to produce a pair of diglists.

The second objective of this demonstration was to have industry geophysicists collaborate with developers of UXO classification methodologies and software. Shaw Environmental geophysicists visited the Vancouver Sky Research office for a one week training session that culminated in the construction and submission of a prioritized diglist based on polarizabilities derived from MetalMapper data.

1.3 REGULATORY DRIVERS

Refer to the Program Office demonstration plan for a discussion of regulatory drivers.

2. TECHNOLOGY

2.1 TECHNOLOGY DESCRIPTION

Magnetic and EM methods represent the main sensor types used for detection of UXO. Over the past 10 years, significant research effort has been focused on developing methods to discriminate between hazardous UXO and non-hazardous scrap metal, shrapnel and geology (e.g. Billings et al., 2010; Bell et al., 2001; Pasion et al., 2007; Tantum et al., 2008; Liao and Carin, 2009). The most promising classification methods typically proceed by first recovering a set of parameters that specify a physics-based model of the object being interrogated. For example, in time-domain electromagnetic (TEM) data, the parameters comprise the object location and the polarization tensor (typically two or three collocated orthogonal dipoles along with their orientation and some parameterization of the time-decay curve). For magnetics, the physics based model is generally a static magnetic dipole. Once the parameters are recovered by inversion, a subset of the parameters is used as feature vectors to guide either a statistical or rule-based classifier.

Magnetic and EM phenomenologies have different strengths and weaknesses. Magnetic data are simpler to collect, are mostly immune to sensor orientation and are better able to detect deeper targets. EM data are sensitive to non-ferrous metals, are better at detecting smaller items and are able to be used in areas with magnetic geology. Only EM data was acquired at Pole Mountain and in the remainder of this report we therefore focus on EM sensing and processing.

There are three key elements that impact the success of the UXO classification process described in the previous paragraphs:

1) Creation of a map of the geophysical sensor data: This includes all actions required to form an estimate of the geophysical quantity in question (i.e. amplitude of EMI response at a given time-channel) at each of the visited locations. The estimated quantity is dependent on the following:
    a. Hardware, including the sensor type, deployment platform, position and orientation system and the data acquisition system used to record and time-stamp the different sensors;
    b. Survey parameters such as line spacing, sampling rate, calibration procedures etc.;

c. Data processing such as merging of position/orientation information with sensor data, noise and background filtering applied;

d. The background environment including geology, vegetation, topography, cultural features, etc.; and

e. Depth and distribution of ordnance and clutter.

2) Anomaly selection and feature extraction: This includes the detection of anomalous regions and the subsequent extraction of a polarization tensor model for each anomaly.

3) Classification of anomalies: The final objective of the demonstration is the production of a dig sheet with a ranked list of anomalies. This will be achieved via statistical classification which will require training data to determine the attributes of the UXO and non-UXO classes.

The focus of this demonstration is on the further testing and validation of the methodologies for 2) and 3) above that have been developed in UXOLab jointly by Sky Research and the University of British Columbia-Geophysical Inversion Facility (UBC-GIF). We now describe each of the three key elements of the technology as identified above.

2.1.1 Creation of a Map of Geophysical Sensor Data

At Pole Mountain, each of the demonstrators provided filtered, background, geo-located geophysical data. No additional pre-processing was applied to the data sets.

2.1.2 Anomaly Selection and Feature Extraction

At this point in the process flow, there is a map of each of the geophysical quantities measured during the survey. The next step in the process is detection of anomalous regions followed by the extraction of features for each of the detected items. For this demonstration, targets have been picked from the EM-61 cart data by the demonstrator, no additional picks were made by Sky/UBC.

In the EMI method, a time varying field illuminates a buried, conductive target. Currents induced in the target then produce a secondary field that is measured at the surface. EM data inversion involves using the secondary field generated by the target for recovery of the position, orientation, and parameters related to the target's material properties and shape. In the UXO community, the inverse problem is simplified by assuming that the secondary field can be accurately approximated as a dipole. In general, TEM sensors use a step off field to illuminate a buried target. The currents induced in the buried target decay with time, generating a decaying secondary field that is measured at the surface. The time-varying secondary magnetic field **B**(*t*) at a location **r** from the dipole **m**(*t*) is computed as:

$$\mathbf{B}(t) = \frac{\mu_o}{4\pi r^3}\mathbf{m}(t)\cdot\left(3\hat{\mathbf{r}}\hat{\mathbf{r}} - \overline{\overline{\mathbf{I}}}\right) \qquad (1)$$

where $\hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|$ is the unit-vector pointing from the dipole to the observation point, **I** is the 3 x 3 identity matrix, μ_o = 4 π x 10^{-7} H/m is the permittivity of free space and $r = |\mathbf{r}|$ is the distance between the center of the object and the observation point.

The dipole induced by the interaction of the primary field $\mathbf{B}_o$ and the buried target is given by:

$$\mathbf{m}(t) = \frac{1}{\mu_O}\overline{\overline{\mathbf{M}}}(t)\cdot \mathbf{B}_O \tag{2}$$

The induced dipole is the projection of the primary field $\mathbf{B}_o$ onto the target's polarizability tensor M(t). The polarizability tensor is assumed to be symmetric and positive definite and so can be decomposed as

$$\boldsymbol{M}(t) = \boldsymbol{A}^T \mathbf{L}(t)\boldsymbol{A} \tag{3}$$

with ***A*** an orthogonal matrix which rotates the coordinate system from geographic coordinates to a local, body centered coordinate system. The diagonal eigenvalue matrix **L**(t) contains the principal polarizabilities L_i(t) (i = 1, 2, 3), which are assumed to be independent of target orientation and location. Features derived from the dipole model, in particular amplitude and decay of the principal polarizabilities, have been successfully used to discriminate between targets of interest and non-hazardous metallic clutter. These parameters are useful because, to first order, a conductor can be modeled as a simple LR loop which is inductively coupled to transmitters and receivers on the surface. The current response of this loop is a decaying exponential which is fully described by an amplitude and time constant (West and Macnae, 1991). The TEM dipole model generalizes this simple circuit model to account for target size and shape. This latter property is represented by the principal polarizabilities, which decay independently in time and are approximately aligned with the semi-major and minor axes of the target.

Equal transverse (secondary and tertiary) polarizabilities indicate an axi-symmetric target. Most ordnance can be treated as bodies of revolution (Shubitidze et al., 2002), and so equality of transverse polarizabilities has been proposed as a useful feature for discriminating between TOI and irregularly-shaped clutter. However, in practice it has been difficult to reliably estimate target shape using data from mono-static, vertical-component sensors conventionally deployed for UXO detection. This is because monostatic data often cannot adequately interrogate the transverse response of buried targets.

Recent advances in TEM sensor technology for UXO detection have helped address these limitations. For example, the MetalMapper sensor is comprised of an array of 7 receivers that measure 3 orthogonal components of the secondary field generated by 3 orthogonal transmitter loops that are fired sequentially. This multi-static, and multi-transmitter configuration provides a very rich data set which is better able to constrain target depth and transverse polarizabilities than a mono-static sensor.

When solving parametric inverse problems, it is usually sufficient to minimize a data norm quantifying the misfit between observed (d^{obs}) and predicted data

$$\phi_d = \left\| V_d^{-1/2}\left(d^{obs} - d^{pred}\right)\right\|^2 \tag{4}$$

with $d^{pred} = F(m)$ generally a nonlinear functional of the model m, and $V_d^{-1/2}$ a (usually diagonal) covariance matrix specifying estimated errors on the data. Bound constraints are also typically imposed to ensure that physically reasonable model parameters are obtained (e.g. polarizabilities should be positive). In the case of TEM data, the model is parameterized in terms of target location and orientation, as well as principal polarizabilities at each time channel. Equation 4 is minimized by first estimating the target location, followed by estimating of polarizabilities at each time channel. Decoupling the time channels in this way makes the inversion less sensitive to the specified uncertainties, but produces

polarizabilities that are less smooth as a function of time.

2.1.3 Classification of Anomalies

At this stage in the process, we have feature vectors for each anomaly and need to decide which items should be excavated as potential UXO. For this demonstration, we employ statistical classification techniques which have proven to be very effective at classification at various test sites (e.g. Billings, 2010). Statistical classifiers have been applied to a wide variety of pattern recognition problems, including optical character recognition, bioinformatics and UXO classification. Within this field there is an important dichotomy between supervised and unsupervised classification. Supervised classification makes classification decisions for a test set comprised of unlabelled feature vectors. The classifier performance is optimized using a training data set for which labels are known. In unsupervised classification there is only a test set; labels are unknown for all feature vectors. Most applications of statistical classification algorithms to UXO discrimination have used supervised classification; the training data set is generated as targets are excavated. More recently, unsupervised methods have been used to generate a training data set that is an informative sample of the test data (Zhang2004). In addition, semi-supervised classifiers, which exploit both labeled data and the topology of unlabelled data, have been applied to UXO classification in one study (Zhang2004).

A popular and proven algorithm employed in the machine learning community for supervised classification and regression problems is the support vector machine (SVM). For the Camp Butner data sets we used this classifier to produce a number of ranked dig sheets, and so here we provide a brief development of the algorithm. We begin with a training set of *M* feature vectors in an *N*-dimensional feature space

$$\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_M\}, \quad x_i \in \boldsymbol{R}^N. \tag{5}$$

In a simple two-class case, the feature vectors have associated labels

$$\boldsymbol{Y} = \{y_1, y_2, \ldots y_M\}, \quad y_i \in \{-1,1\}. \tag{6}$$

Define a support plane for a given class to be a line (or hyperplane in higher dimensions) so that all feature vectors in that class fall to one side of the support plane, as illustrated in Figure 1. The margin between classes is the perpendicular distance between support planes. The SVM algorithm then tries to find an optimal decision plane by maximizing the margin, subject to the constraint that the data are classified correctly. Referring to figure 1, the decision function is

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b \tag{7}$$

with $\boldsymbol{w}$ the normal vector, and *b* an offset. The decision boundary is defined by points for which $f(\boldsymbol{x}) = 0$, and a ranking of feature vectors can be obtained by thresholding on the decision function.

To maximize the margin we then find that we can minimize $\|\mathbf{w}\|^2/2$. It can be shown that this optimization problem can be solved by maximizing the dual problem

$$\mathbf{L}_d = \left[\sum_{i=1}^{M} \alpha_i - \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i y_i \left(\mathbf{x}_i^T \mathbf{x}_j\right) y_j \alpha_j\right] \tag{8}$$

With the constraints

$$\sum_{i=1}^{M} \alpha_i y_i = 0.$$

$$\alpha_i \geq 0, \forall i. \tag{9}$$

The dual problem gives us a solution for the normal to the hyperplane

$$\boldsymbol{w} = \sum_{i=1}^{M} \alpha_i y_i \, \mathbf{x}_i. \tag{10}$$

The solution has the property that the majority of coefficients α_i are zero, with nonzero values corresponding to feature vectors lying on the support planes (these are termed the support vectors).

A more general formulation of the SVM allows for nonlinear decision boundaries with overlapping classes. The idea is to map the feature data to a higher-dimensional space where the data become separable and to construct the optimal separating hyperplane in this space. In equation 8 we need only evaluate the inner product of the feature vectors. Hence to construct the decision boundary in another space we need only compute inner products in that space. The optimization problem becomes

$$\mathbf{L}_d = \left[\sum_{i=1}^{M} \alpha_i - \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i y_i K\left(\mathbf{x}_i, \mathbf{x}_j\right) y_j \alpha_j\right] \tag{11}$$

with the same constraint as before (equation 9). The most common choices of kernel functions $K(\mathbf{x}_i, \mathbf{x}_j)$ for nonlinear SVMs are radial basis functions and polynomials. While the choice of kernel functions is often arbitrary, practical experience has shown that SVMs trained with different kernels often have a large percentage of support vectors in common.

Kernel function	$K(\mathbf{x}_i, \mathbf{x}_j)$
Radial Basis function	$\exp\left(-\frac{\|\mathrm{x}_i - \mathrm{x}_j\|^2}{2\sigma^2}\right)$
Degree p polynomial	$\left(x_i \cdot x_j \ + 1\right)^p$

Table 1: *Kernel functions commonly used with nonlinear support vector machines.*

The complexity of the decision boundary is controlled by the kernel parameter (σ in the case of the radial basis function). A small kernel parameter results in a decision boundary that fits the training data closely, but may not generalize well to the test data.

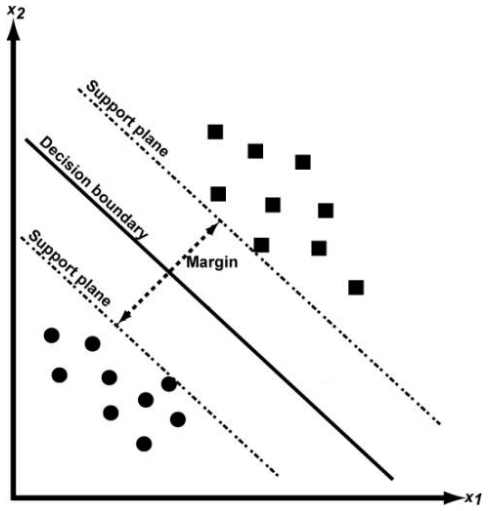


Figure 1. *SVM formulation for constructing a decision boundary. The decision boundary bisects support planes bounding the classes.*

2.1.4 Sky Research Classification Software Suite: UXOLab

The methodologies for data processing, feature extraction, and statistical classification described above have been implemented within the UXOLab software environment, which was used for this demonstration. UXOLab contains modules for data visualization, data inversion, quality control of inversion results, training data selection, and diglist creation via statistical or rule-based classification strategies. UXOLab is a Matlab based software package developed over a six year period at the UBC-GIF, principally through funding by the United States Army Corps of Engineers-Engineering Research and Development Center (USACE-ERDC) (DAAD19-00-1-0120). Over the past five years, Sky Research and UBC-GIF have considerably expanded the capabilities of the software.

2.2 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY

The main advantage of the technology is a potential reduction in the number of non-hazardous items that need to be excavated, thus reducing the costs of UXO remediation. Advantages of UXOLab and the algorithms within the package include:

- All the functionality required to process raw geophysical data, detect anomalous regions, and perform geophysical inversion and classification.
- Extensive set of algorithms for rule-based (e.g. polarizability match, decay characteristics, polarizability amplitude, etc.) and statistical classification (e.g. probabilistic neural network, support vector machine and relevance vector machine, etc.) algorithms.
- Configuration in a modular fashion, so that as new sensor technologies become available (e.g. new TEM systems with multi-component receivers etc), the inversion functionality will be immediately available to those new sensor systems.

The principal disadvantage is that UXOLab is written in Matlab and has not been configured for general use by contractors and non-specialists. However, as part of ESTCP MR-201004 we are presently working on transitioning our inversion algorithms to an API that will be generally accessible.

3. PERFORMANCE OBJECTIVES

The performance objectives for this demonstration are summarized in Table 2. There are objectives for both the data collection and data analysis demonstrators.

***Table 2**. Performance Objectives for This Demonstration*

Performance criterion				EM-61	MetalMapper	
Performance Objective	**Metric**	**Data Required**	**Success Criteria**	**Statistical**	**Statistical**	**Library**
Maximize correct classification of munitions	Number of targets-of-interest retained	• Prioritized anomaly lists • Scoring reports from (IDA)	Approach correctly classifies all TOI	Yes	Yes	Yes
Maximize correct classification of non-munitions	Number of false alarms eliminated	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of FA > 30% while retaining all TOI	Yes	Yr 1: Yes Yr 2:Yes	Yr 1: Yes Yr 2: Yes
Specification of no-dig threshold	Pd of correct classification and #FA at operating point	• Demonstrator-specified threshold • Scoring reports from IDA	Threshold achieves criteria above	Yes	Yr 1: Yes Yr 2:Yes	Yr 1: Yes Yr 2:Yes
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze"	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies	Yes	Yr 1: Yes Yr 2:Yes	Yr 1: Yes Yr 2:Yes
Correct estimation of target parameters (positions)	Accuracy of estimated target parameters	• Demonstrator target parameters • Results of intrusive investigation	X, Y < 15 cm (1σ) Z < 10 cm (1σ)	Not Calcula-ted	Yes (13.2cm, 13.9cm) Yes (5.1 cm)	Yes (13.2cm, 13.9cm) Yes (5.1 cm)

The first three analysis objectives refer to the classification part of the demonstration with the first two referring to the best results from each approach in a retrospective analysis and the third addressing how well each demonstrator is able to specify the correct threshold in advance. The final two objectives refer to the feature extraction part of the demonstration.

OBJECTIVE: MAXIMIZE CORRECT CLASSIFICATION OF MUNITIONS

This is one of the two primary measures of the effectiveness of the classification approach. By collecting high quality data and analyzing those data with advanced parameter estimation and classification algorithms we expect to be able to classify the targets with high efficiency. This objective concerns the component of the classification problem that involves correct classification of items-of-interest.

Metric: The metric for this objective is the number of items on the master anomaly list that can be correctly classified as munitions by each classification approach.

Data Requirements: Each demonstrator will prepare a prioritized dig list for the targets on the master anomaly list. IDA personnel will use their scoring algorithms to assess the results.

Success Criteria: The objective will be considered to be met if all of the items of interest are correctly labeled as munitions on the prioritized anomaly list.

OBJECTIVE: MAXIMIZE CORRECT CLASSIFICATION OF NON-MUNITIONS

This is the second of the two primary measures of the effectiveness of the classification approach. By collecting high-quality data and analyzing those data with advanced parameter estimation and classification algorithms we expect to be able to classify the targets with high efficiency. This objective concerns the component of the classification problem that involves false alarm reduction.

Metric: The metric for this objective is the number of items-of-interest on the master dig list that can be correctly classified as non-munitions by each classification approach.

Data Requirements: Each demonstrator will prepare a prioritized dig list for the targets on the master anomaly list. IDA personnel will use their scoring algorithms to assess the results.

Success Criteria: The objective will be considered to be met if more than 30% of the non-munitions items can be correctly labeled as non-munitions while retaining all of the targets-of-interest on the dig list.

OBJECTIVE: SPECIFICATION OF NO-DIG THRESHOLD

In a retrospective analysis as will be performed in this demonstration, it is possible to tell the true classification capabilities of a classification procedure based solely on the prioritized dig list submitted by each demonstrator. In a real-world scenario, all targets may not be dug so the success of the approach will depend on the ability of an analyst to accurately specify their dig/no-dig threshold.

Metric: The probability of correct classification, P_{class}, and number of false alarms, N_{fa}, at the demonstrator-specified threshold are the metrics for this objective.

Data Requirements: Each demonstrator will prepare a ranked anomaly list with a dig/no-dig threshold indicated. IDA personnel will use their scoring algorithms to assess the results.

Success Criteria: The objective will be considered to be met if more than 30% of the non-munitions items can be correctly labeled as non-munitions while retaining all of the targets-of-interest at the demonstrator-specified threshold.

OBJECTIVE: MINIMIZE NUMBER OF ANOMALIES THAT CANNOT BE ANALYZED

Anomalies for which reliable parameters cannot be estimated cannot be classified by the classifier. These anomalies must be placed in the dig category and reduce the effectiveness of the classification process.

Metric: The number of anomalies for which reliable parameters cannot be estimated is the metric for this objective.

Data Requirements: Each demonstrator that estimates target parameters will provide a list of all parameters as part of their results submission along with a list of those anomalies for which parameters could not be reliable estimated.

Success Criteria: The objective will be considered to be met if reliable parameters can be estimated for > 90% of the anomalies on each sensor anomaly list.

OBJECTIVE: CORRECT ESTIMATION OF TARGET PARAMETERS

This objective involves the accuracy of the target parameters that are estimated in the first phase of the analysis. Successful classification is only possible if the input features are internally consistent. The obvious way to satisfy this condition is to estimate the various target parameters accurately.

Metric: Accuracy of estimation of target parameters is the metric for this objective.

Data Requirements: Each demonstrator that estimates target parameters will provide a list of all parameters as part of their results submission. IDA analysts will compare these estimated parameters to those measured during the intrusive investigation and determined via subsequent in-air measurements.

Success Criteria: The objective will be considered to be met if the estimated βs are within ± 20%, the estimated X, Y locations are within 15 cm (1σ), the estimated depths are within 10 cm (1σ), and the estimated size is within ± 20%.

4. SITE DESCRIPTION

The Pole Mountain Target and Maneuver Area (PMTMA) is a 62,448.15 acre site located in Laramie, WY. The Pole Mountain demonstration will be conducted in the Bisbee Hill Maneuver Area. See the Program Office demonstration plan for more details on the site.

4.1 SITE SELECTION

This site was chosen as the next in a progression of increasingly more complex sites for demonstration of the classification process. The first site in the series, Camp Sibert, was open field survey conditions with only one target-of-interest and item “size” was an effective discriminant. At this site, there was a

wider mixture of munitions types, with the smallest target of interest being a small ISO. The site is open and relatively flat, with trees and topography not an issue.

4.2 SITE HISTORY

See the Program Office demonstration plan.

4.3 SITE GEOLOGY

See the Program Office demonstration plan.

4.4 MUNITIONS CONTAMINATION

See the Program Office demonstration plan.

5. TEST DESIGN

See the Program Office demonstration plan for a description of the test design for the overall project.

Sky Research/UBC-GIF processed data and delivered the following digsheets:

1) Geonics EM-61: Statistical classification of features derived from the Geonics EM-61 data and the production of a ranked dig sheet;
2) MetalMapper Cued Interrogation Data
    a. Statistical Classification: Statistical classification via a support vector machine (SVM) classifier applied to dipole polarizabilities derived from single and two source inversion of MetalMapper anomalies.
    b. Rule Based Classification: Classification achieved by comparing dipole polarizabilities derived from single and two source inversion of MetalMapper anomalies, to polarizabilities contained with a library of reference items. A pair of diglists – based on aggressive and less-aggressive version of the classifier – was submitted.

Shaw Environmental, with assistance and guidance from Sky Research data processors, submitted a library match based digsheet

6. DATA ANALYSIS AND RESULTS

6.1 EM61 diglists

To generate ordered diglists from Pole Mountain EM-61 data, we used the *decay* parameter estimated from dipole model fits

$$\boldsymbol{decay} = \frac{\boldsymbol{L_{tot(t_4)}}}{\boldsymbol{L_{tot(t_1)}}} \tag{12}$$

with L_{tot} the total polarizability. Because target size is poorly constrained by EM-61 data (owing to poor constraints on depth), we do not use a size parameter derived from the dipole model. We instead use the maximum predicted amplitude of the data at the first time channel as a proxy for target size. Data amplitude is useful for identifying large TOI (e.g. 75 mm) in the early stages of digging, and so training in a data amplitude/decay feature space can provide some initial improvement in discrimination performance relative to a strategy that relies on decay alone.

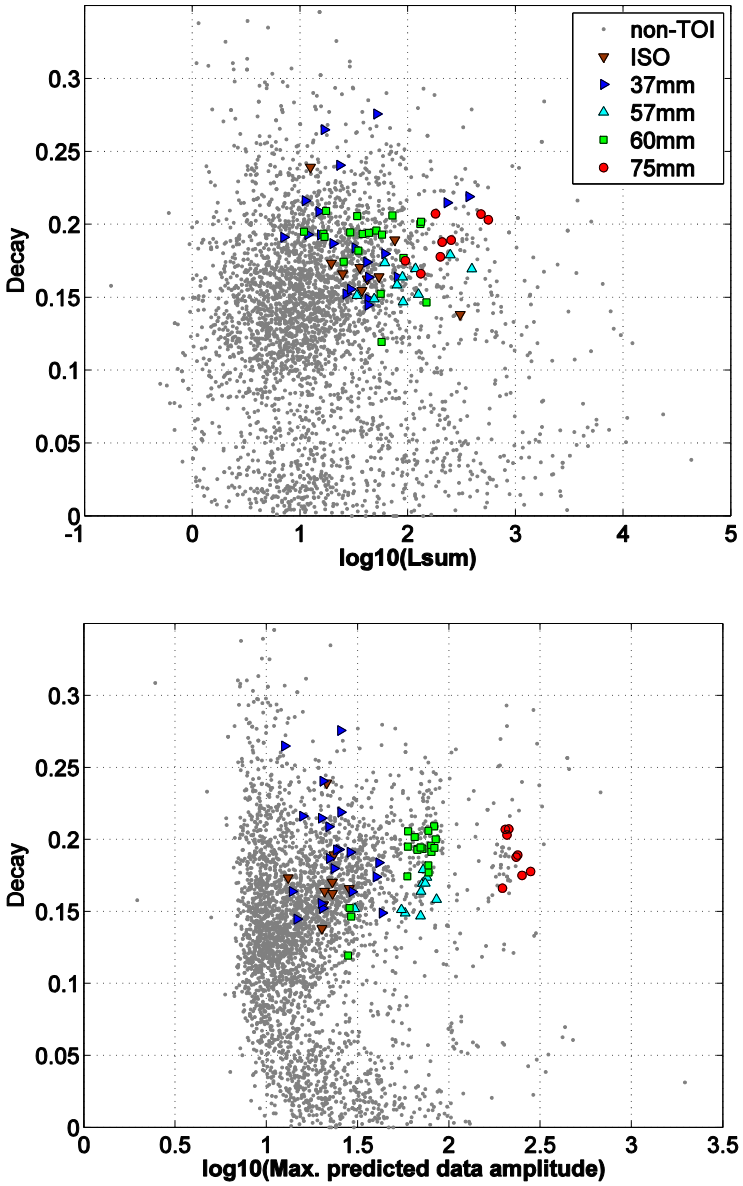


Figure 2. *Comparison of polarizability size (top) and data amplitude (bottom) features for EM-61 data, Pole Mt. Year 1.*

This is illustrated in Figure 2 for the Pole Mt. year 1 EM-61 data: large ordnance are better clustered in data amplitude vs. polarizability size. However, for identification of ISO and 37mm targets the polarizability size parameter appears to have some advantage.

To classify the EM-61 data, we trained a probabilistic neural network (PNN) classifier in a two-dimensional data amplitude/decay feature space. The training distribution of TOI used feature vectors from the previous ESTCP demonstration at Camp Butner. A similar size range of TOI was expected at Pole Mountain, and there is a good correspondence between the training TOI and large amplitude, slow-decaying feature vectors in the test data (Figure 3). For training non-TOI we use unlabeled test items with the maximum misfit to known TOI as *assumed* non-TOI. This technique is described in more detail on page 36.

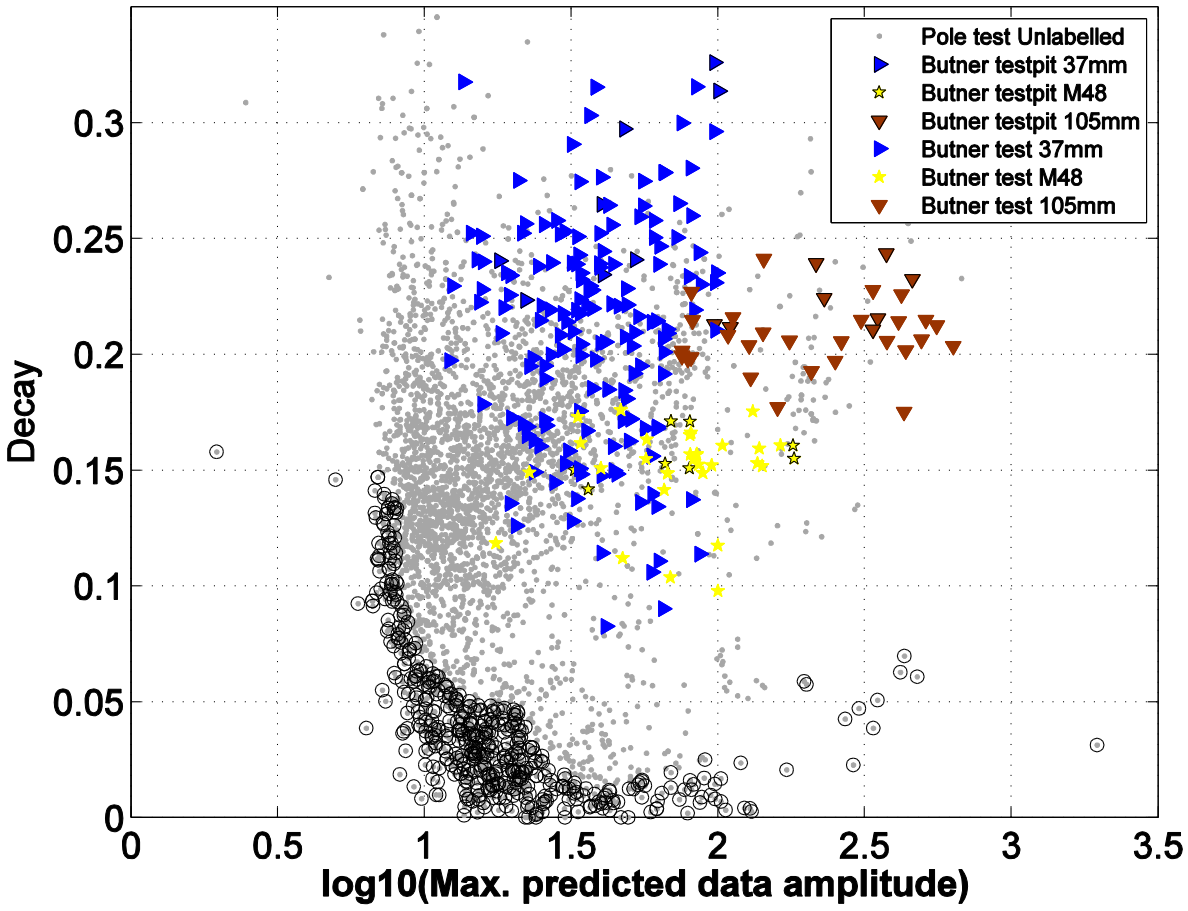


Figure 3. *Training TOI for EM-61 data (colored markers). Training non-TOI are shown as black circles.*

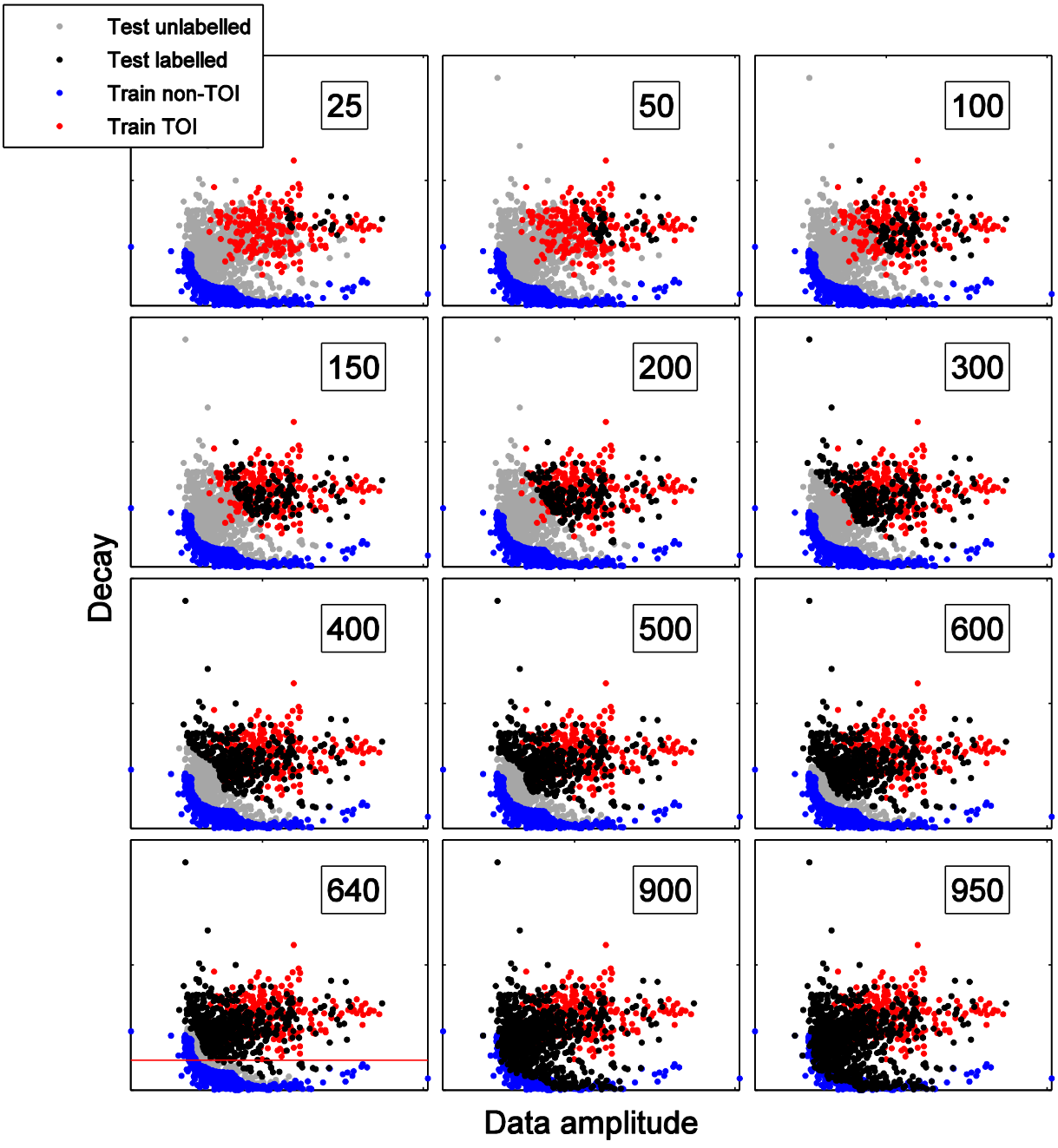


Figure 4. *Dig order for PNN classifier applied to Peale Mountain Year 1 EM-61 data. The number in the top right of each subplot indicates the first N labeled test items found during digging, displayed as black markers in the test data. Digging in the first stage was terminated after 1200 items. Horizontal red line indicates a decay rate cut-off of 0.08.*

Figure 4 shows the resulting dig order for a probabilistic neural network (PNN) classifier applied to the Pole Mountain Year 1 test data. This classifier was trained in a two-dimensional data amplitude/decay feature space. The resulting ROC for the classification with the EM-61 is mediocre (Figure 5).

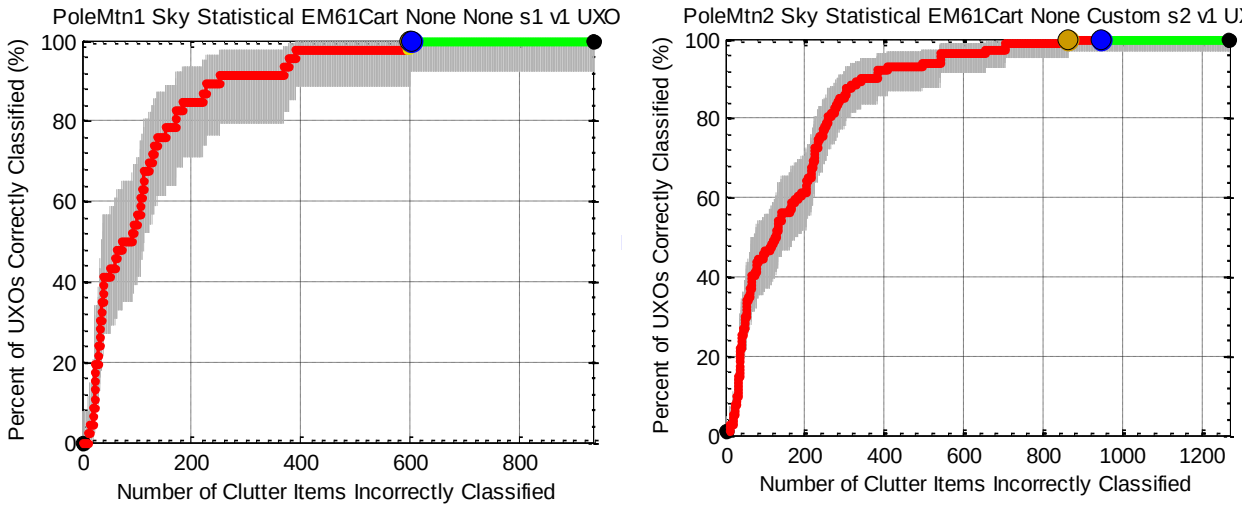


Figure 5. *ROC for classification of Pole Mountain Year 1 and Year 2 EM-61 data.*

In Figure 5 we remark that the last TOI was placed in the "can't decide" dig category and so appears very close to the selected stop dig point. This item had a very poor data fit and should perhaps have been labeled as "can't analyze." This would have significantly reduced the false alarm rate at which all TOI were identified to around 400 clutter items. Some incremental improvement might also be obtained in this case by training on polarizability size (rather than data amplitude), but the MetalMapper performance is so clearly superior (as described in the following sections) that we conclude that EM-61 discrimination is not worthwhile at this site. We note that the EM61 data would have some limited use as a pre-screener. By using the EM61 data, the number of anomalies visited by the MetalMapper could be reduced by approximately 25%.

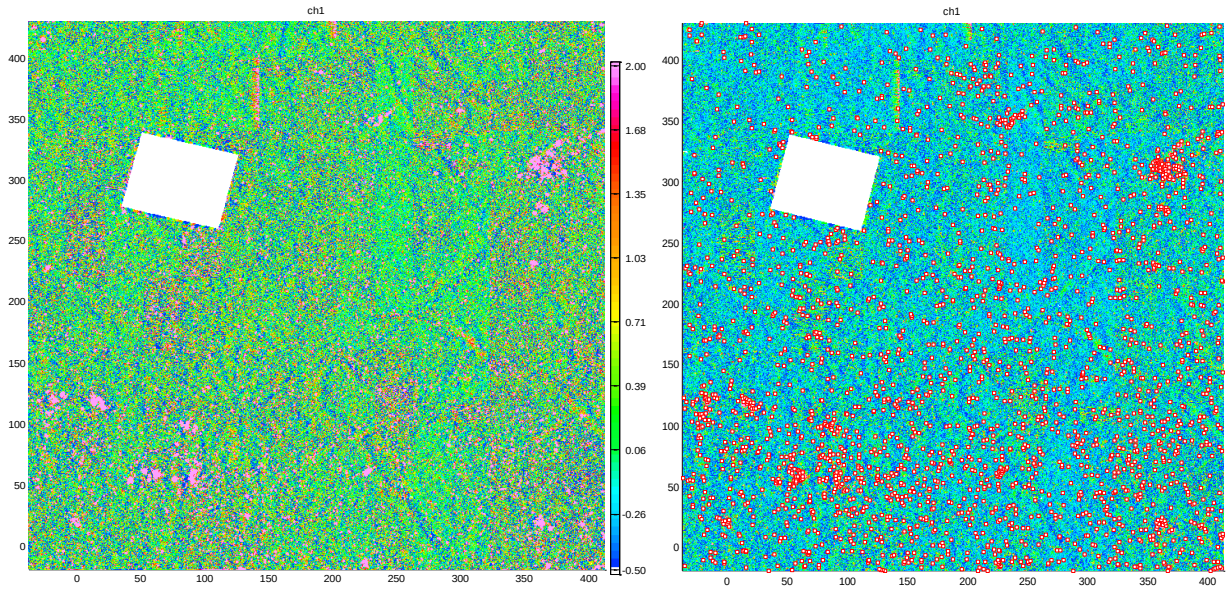
6.2 MetalMapper static feature extraction and discrimination

6.2.1 Feature extraction

2370 anomalies were divided into two groups (Figure 6) by the Program Office to simulate a two-year field program. Year 1 comprised the southern 986 anomalies; year 2 comprised the northern 1384 anomalies. A final dig list for the year 1 data was submitted before the year 2 data were received for analysis. Complete ground truth for year 1 was not available until after both final dig lists were submitted. (i.e., we did not have full ground truth for year 1 when we started analysis of year 2 data). Analysis of the year 2 data followed the same procedure used for the year 1 data, thus we describe in detail only the analysis of the year 1 data.

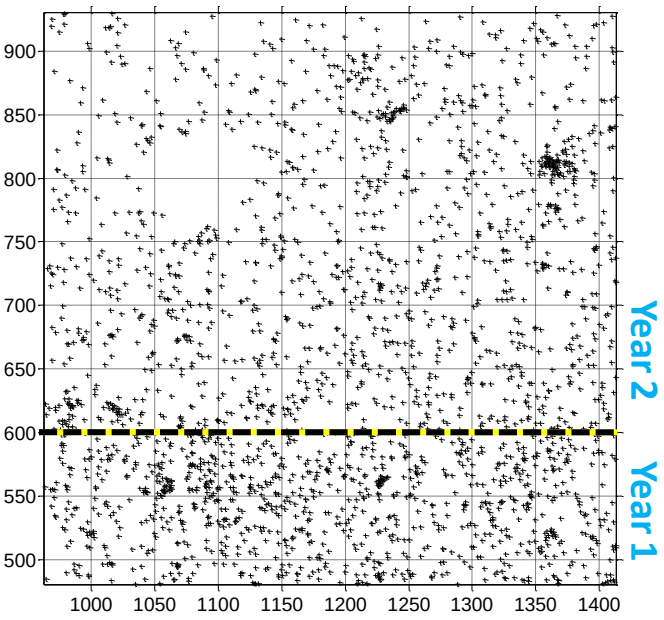
Background-corrected MetalMapper cued data for the 986 year 1 anomalies were received as a set of CSV files. The data were inverted in UXOLab using a sequential inversion approach to estimate target location, depth and primary polarizabilities. Instrument height above the ground was assumed to be 7 cm. Noise standard deviation estimates were not available, so a constant noise value of 1 over all time channels was used. Target location was constrained to lie between ±0.5 m in both X and Y directions relative to the center of the MetalMapper. Target depth was constrained to lie between –1.2 and 0 m. The initial optimization for target location identified up to three starting models to input into the subsequent estimation of polarizabilities. The data for each anomaly were inverted using both a single-object inversion (SI) and multi-object (MI, i.e. two-object) inversion.

Visual QC of the data was performed using a newly developed QC tool which provides an overview of the observed and predicted data, predicted model parameters, and measures of data/model quality. Predicted polarizabilities were compared to reference polarizabilities for various TOI derived from test pit and IVS measurements. The library of reference items was augmented based on ground truth obtained through training data requests. During QC the primary objectives were to pass the best model (i.e., the one with the most UXO-like polarizabilities), flag high-likelihood TOI, and fail unlikely models. The latter most commonly applied to one of the MI models which frequently are unrealistic (e.g., deep, large in magnitude, and sometimes located on or near a horizontal inversion boundary; e.g., Figure 7) but sometimes provide the best fit to the reference polarizabilities. In some cases the decision on which model to pass was not obvious. In these cases more than one model was passed; the classification procedure would consider all passed models and effectively use the one that is most like TOI.



(a) Channel 1 of Geonics EM61data

(b) Anomaly locations overlaid on EM61 data



(c) Anomaly map showing division of 2370 anomalies to simulate a 2-year field program. The number of anomalies for year 1 and year 2, respectively, was 986 and 1384.

Figure 6. *Geonics EM61 data and anomaly map.*

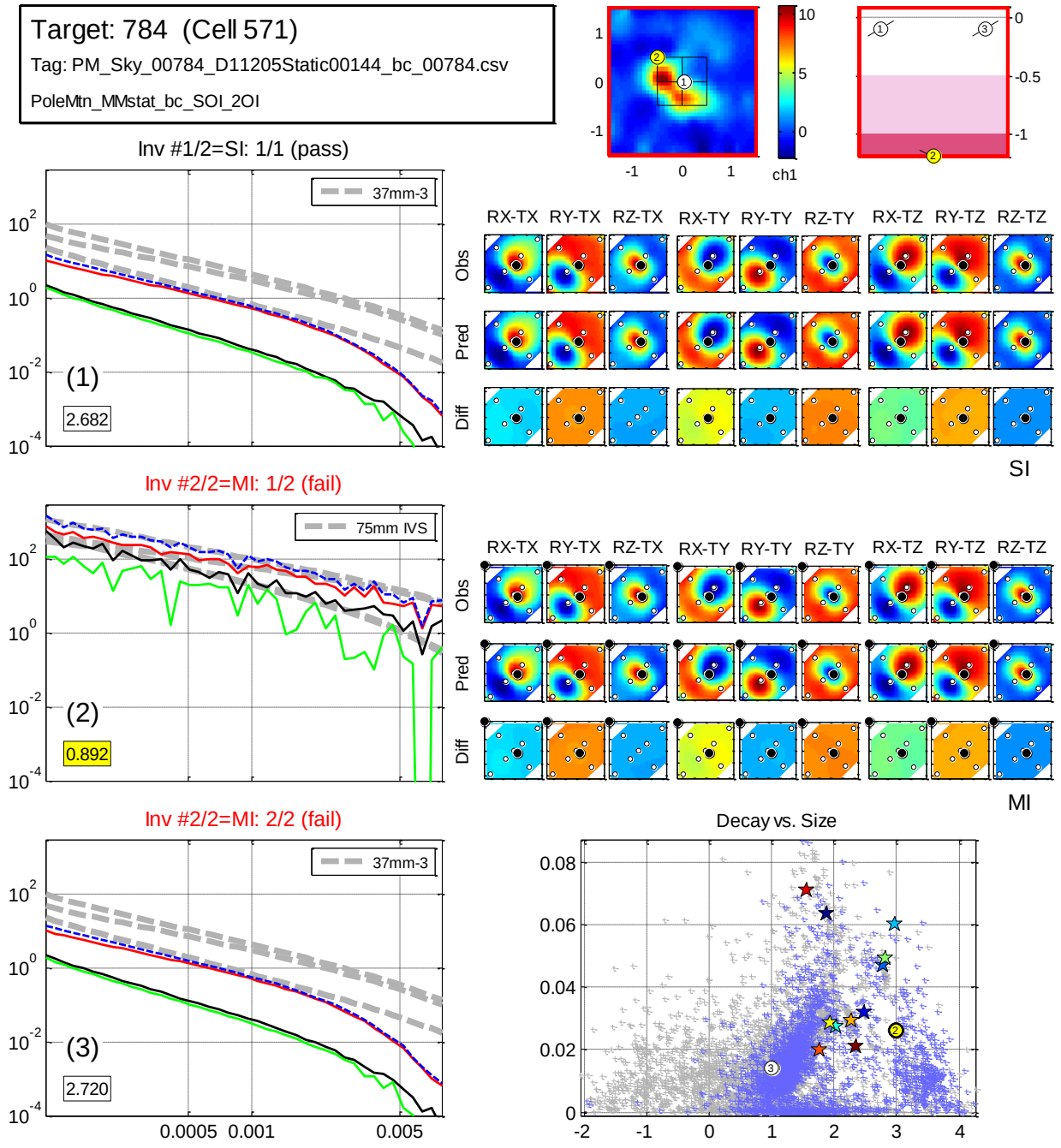


Figure 7. *Example of an unrealistic MI model (anomaly 784; scrap). The first model of the MI (model 2) provides (by far) the best fit to the reference polarizabilities (misfit = 0.892), but the predicted depth of 1.2m, location at the corner of the instrument, and high amplitude of the polarizabilities, especially in relation to the relatively weak EM61 anomaly, are indicators that this model is an artifact of the MI process. Accordingly, this model was failed during QC.*

6.2.2 Discrimination

6.2.2.1 Training data selection

Figure 8 shows the distribution of all passed year 1 models in decay-size feature space.

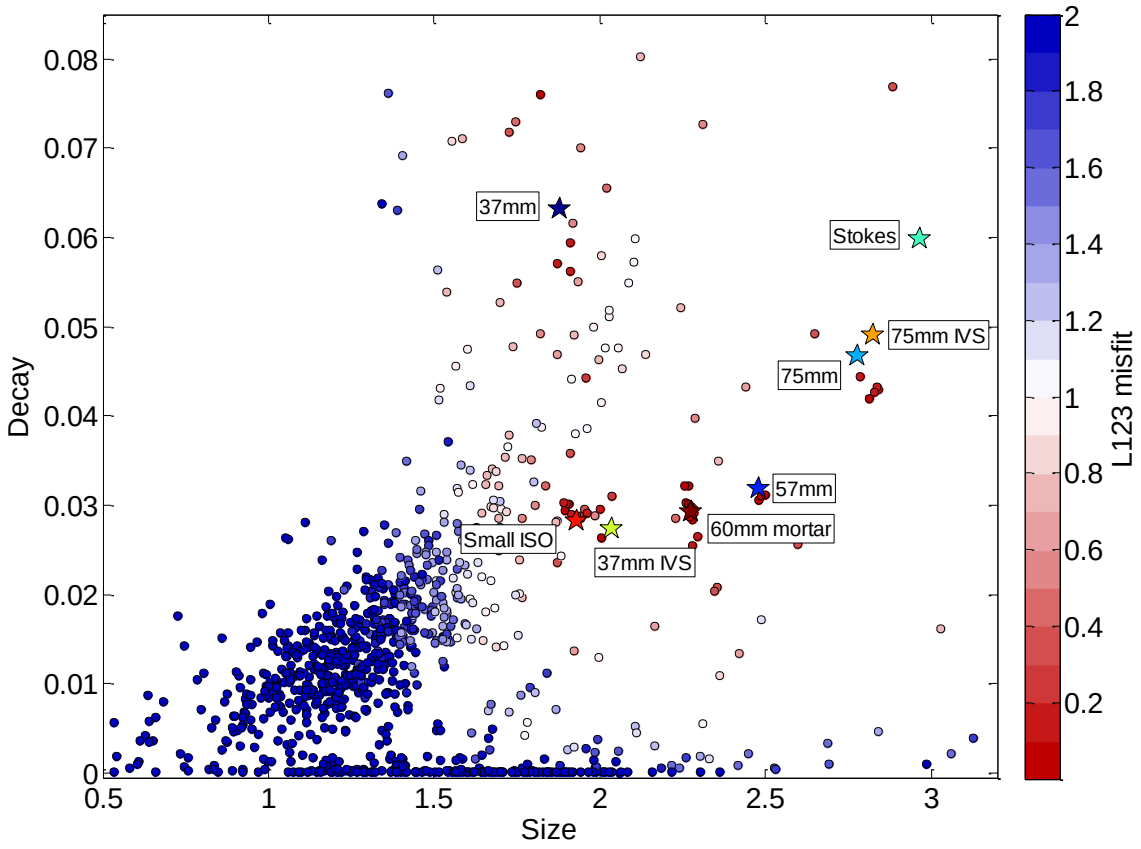


Figure 8.*Distribution of all year 1 passed models in decay(t1,t29) versus size(t1) feature space, where size(t1) is the total polarizability measured at the first time channel (t1=0.106ms), and decay(t1,t29) = size(t1)/size(t29) where t29=2.006ms. Features (dots) are color-coded by polarizability misfit with best fitting reference item. Labeled stars show location of library reference items in feature space.*

A newly developed training data selection tool was used to assist with the selection of training data. This tool is used to find clusters of items with similar polarizabilities. A two dimensional feature space defined by the analyst (e.g. size/decay, primary/secondary polarizabilities) is used to provide a visual guide for manually selecting anomalies. Training data were not requested for clusters of items whose polarizabilities closely matched those of known reference items. A basic example of the usage of the training data selection tool is shown in Figures 8 and 9. In this example, a cluster of items sitting between the small ISO and 57mm projectile in feature space turned out to correspond to 60mm mortar projectiles.

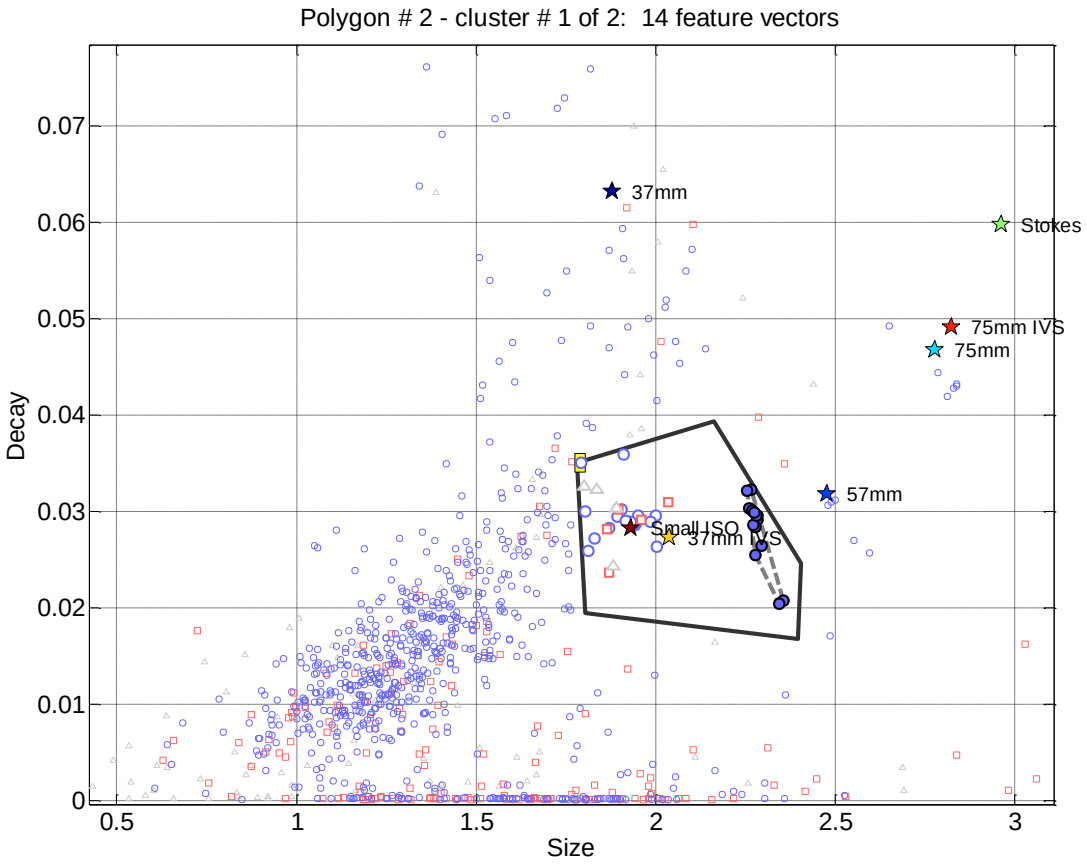


Figure 9. *Example of use of training data selection tool. A polygon (heavy black line) is drawn in feature space. Clusters of items with self-similar polarizabilities within the polygon are automatically found. In this case several clusters were found; one is visible (solid feature symbols surrounded by broken grey line). Polarizabilities for this cluster are shown in Figure 10.*

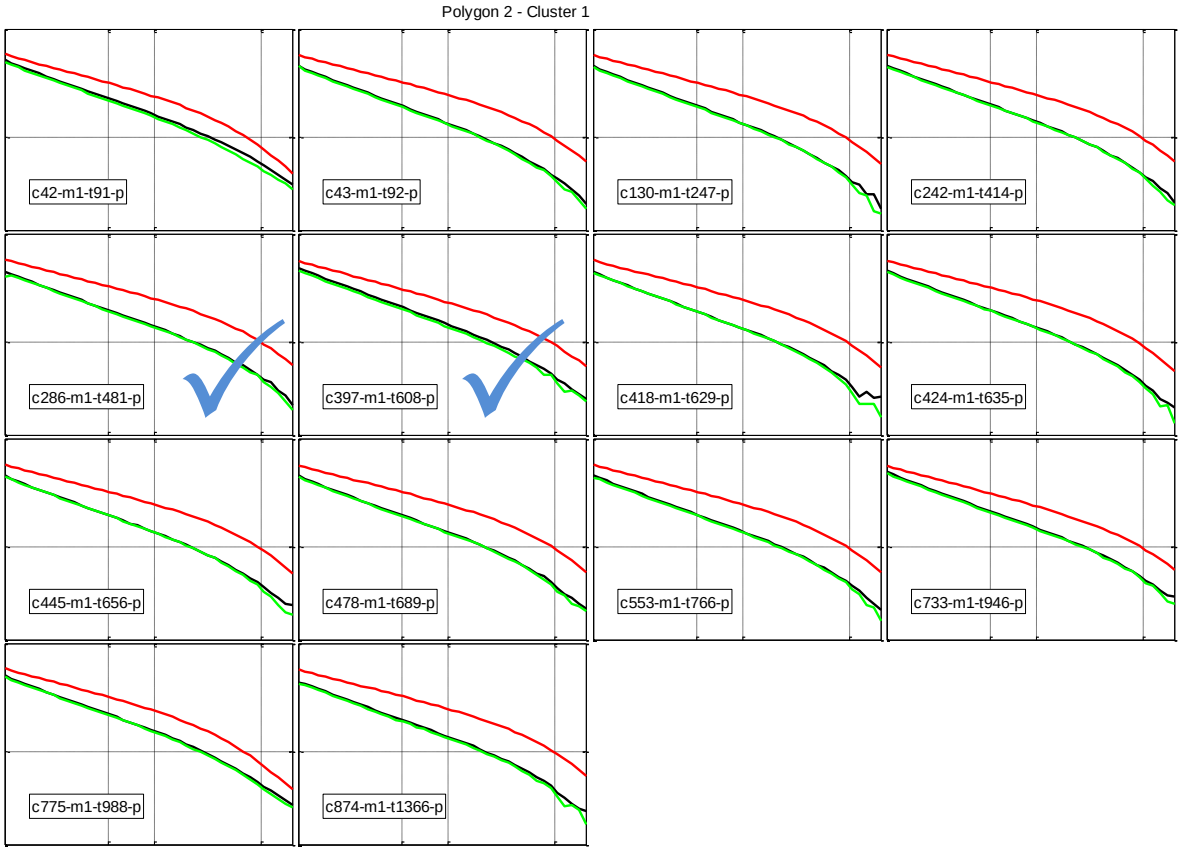


Figure 10. *Polarizabilities for the cluster shown in Figure 9. Training data were requested for two anomalies: 481 and 608 (check marked); ground truth revealed that both of these anomalies correspond to 60mm mortar projectiles.*

Of primary concern was the large group of small, fast-decaying features (black dashed line in Figure 11), which could conceal small, unknown TOI. The polarizabilities of many of the items in this group are UXO-like. We requested ground truth for two of the items in the small cluster shown in Figure 11; both of these turned out to be scrap (light frag). Training data from other small items within the larger group also turned out to be scrap.

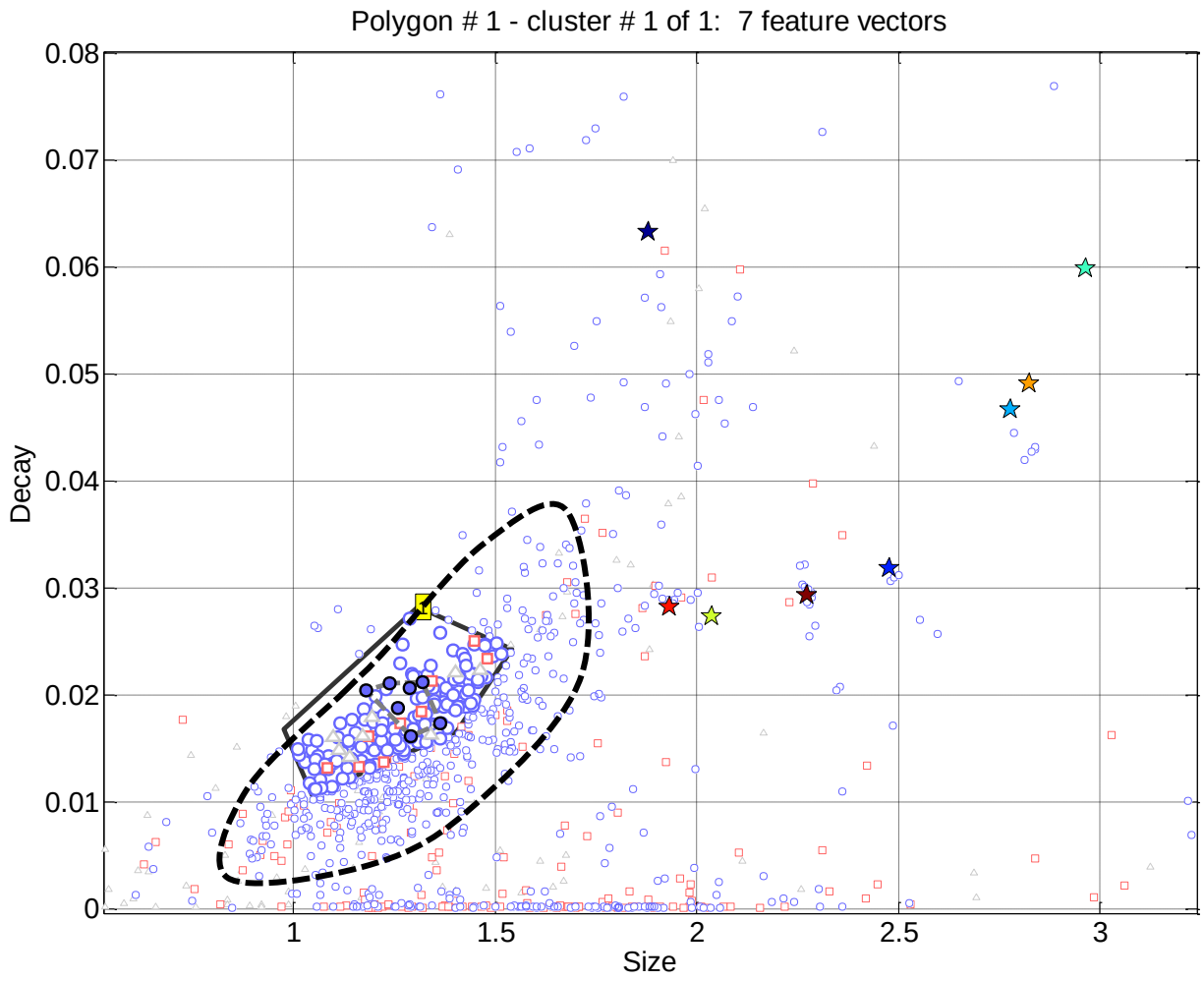


Figure 11. *Dashed line shows large group of small, relatively fast decaying features. Many of these have UXO-like (i.e. equal secondary and smoothly decay) polarizabilities. The polarizabilities of the small cluster of items shown (solid feature symbols surrounded by broken grey line) are shown in Figure 12. Ground truth revealed that these corresponded to scrap.*

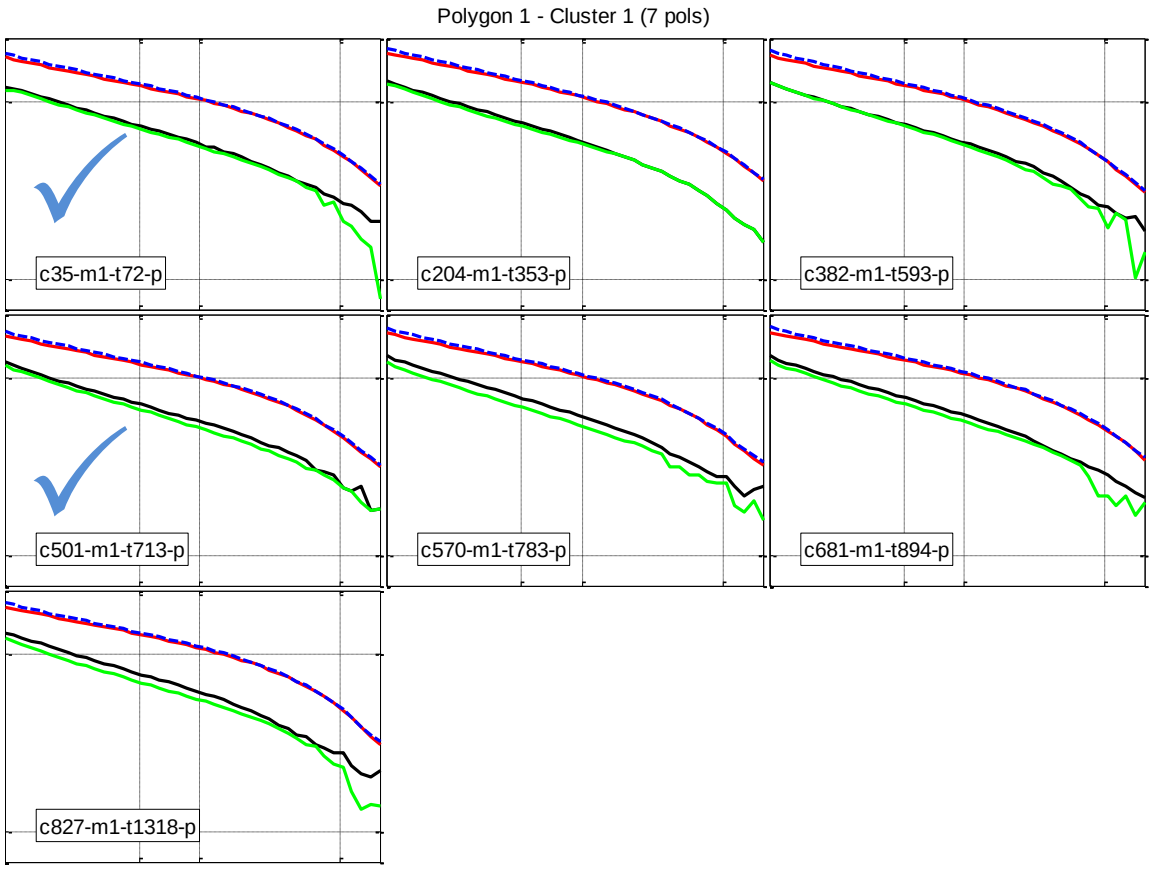


Figure 12. *Polarizabilities for the cluster shown in Figure 11. Training data were requested for two anomalies: 72 and 713 (check marked); ground truth revealed that both of these anomalies correspond to scrap (light frag).*

A large cluster of items with fast decaying polarizabilities and observed data at early times is shown in Figure 13. The polarizabilities, though remarkably consistent, are not UXO-like. Ground truth from training data showed that these items correspond to wires.

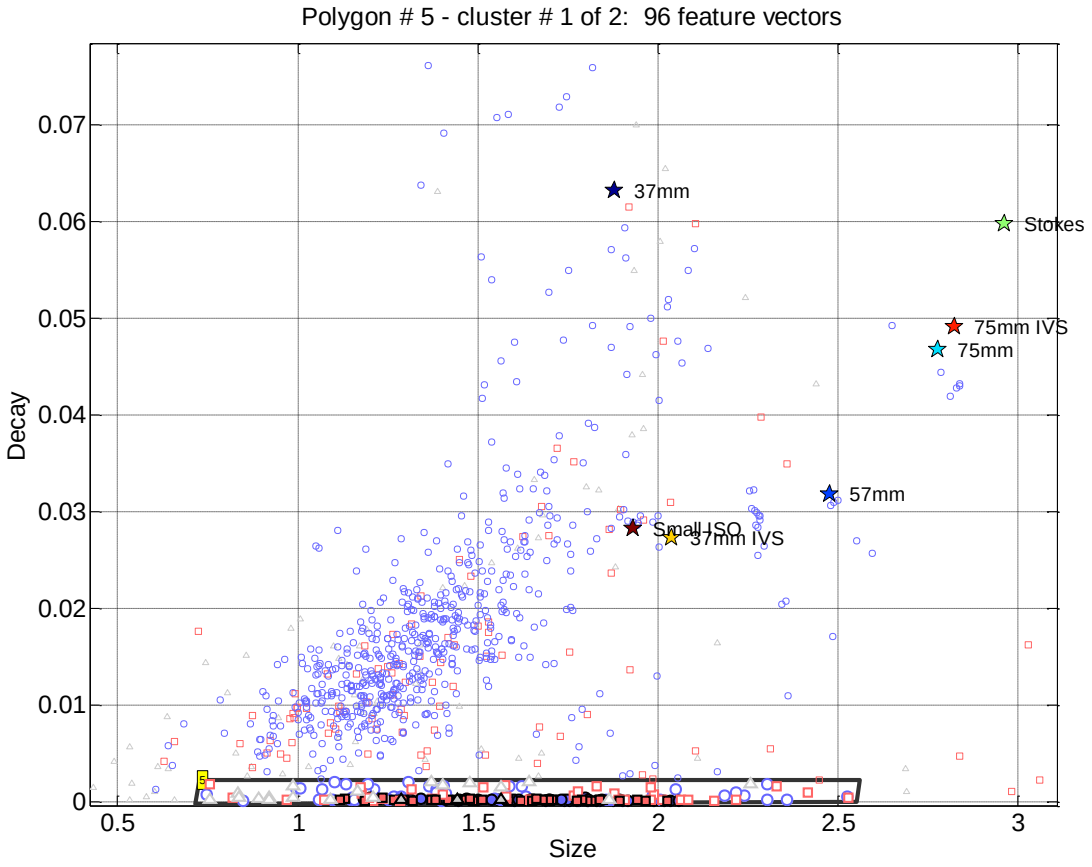


Figure 13. *Cluster with fast decaying polarizabilities at early times. Items in the cluster are solid feature symbols surrounded by broken grey line (the latter is not visible). Polarizabilities for this cluster are shown in Figure 14*

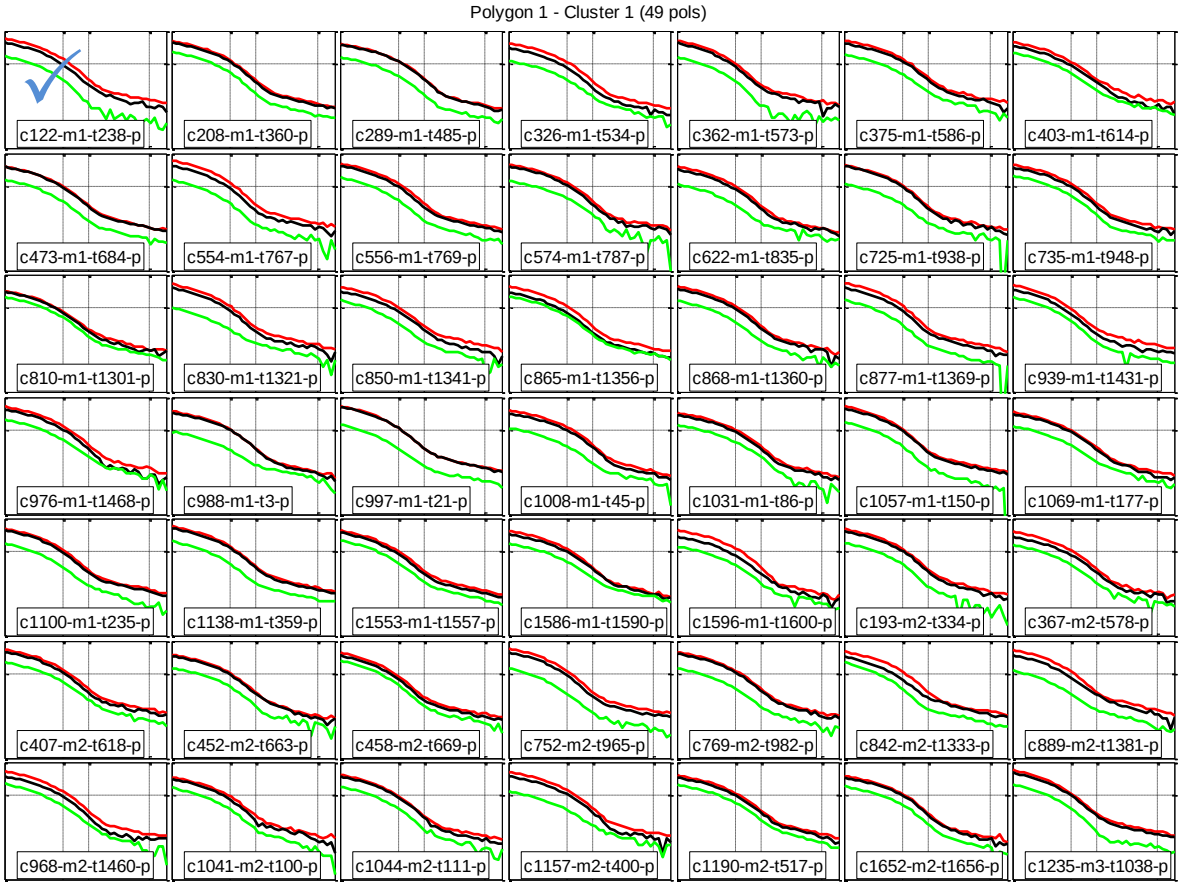


Figure 14. *Polarizabilities for the cluster shown in Figure 13. Training data were requested for Anomaly 238 (check marked); ground truth for this item, and the final ground truth, revealed that all of these items correspond to wires.*

We made three separate requests for a total of 17 training data (Table 3). Of these, four were TOI.

Table 3. *Year 1 training data requests. Items highlighted in yellow are TOI.*

Anomaly_ID	Depth	Dip Angle	Azimuth	Identification	Length	Dig Type
PM1-122	4	0	85	Horseshoe	15	CD
PM1-153	4	0	290	Frag (medium)	14	MD
PM1-238	0	0	0	Wire	25	CD
PM1-439	1	5	90	Frag (light)	17	MD
PM1-439	1	5	90	Nail	7	CD
PM1-481	37	-90	300	60mm mortar	13	TOI
PM1-505	14.5	60	0	Small ISO Item	10	TOI
PM1-569	4	0	180	Horseshoe	11	CD
PM1-572	0	0	0	Small ISO	10	TOI
PM1-608	38	0	130	60mm mortar	13	TOI
PM1-614	0	0	220	Wire	21	CD
PM1-1317	11	20	300	Frag (medium)	9	MD
PM1-1452	4	0	0	Horseshoe	13	CD
PM1-6	6	0	180	Frag (medium)	13	MD
PM1-294	6	0	110	Frag (light)	8	MD
PM1-782	2	0	270	Frag (medium)	8	MD
PM1-72	2	0	160	Frag (light)	8	MD
PM1-713	5	0	40	Frag (light)	7.5	MD

6.2.2.2 Discrimination method

A new dig list tool (Figure 15) was used to determine the digging order. For each anomaly a score S_i is calculated based on a weighted combination of the following parameters

- size of the secondary and tertiary polarizabilities
- minimum misfit with all library reference
- measure of the decay of the total polarizability
- size of the total polarizability
- minimum misfit with a set of "non-ordnance" reference polarizabilities (e.g., horseshoe);
- polarizability quality

Polarizability quality is an *ad hoc* measure of (1) how much the polarizabilities look like those of a typical UXO (i.e., polarizabilities of an axi-symmetric body); and (2) the smoothness of the polarizabilities. We submitted two dig lists: (1) an "aggressive" dig list which included polarizability quality and used all three polarizabilities to calculate misfit; and (2) a "not aggressive" list which did not include polarizability quality and used total polarizability for misfit.

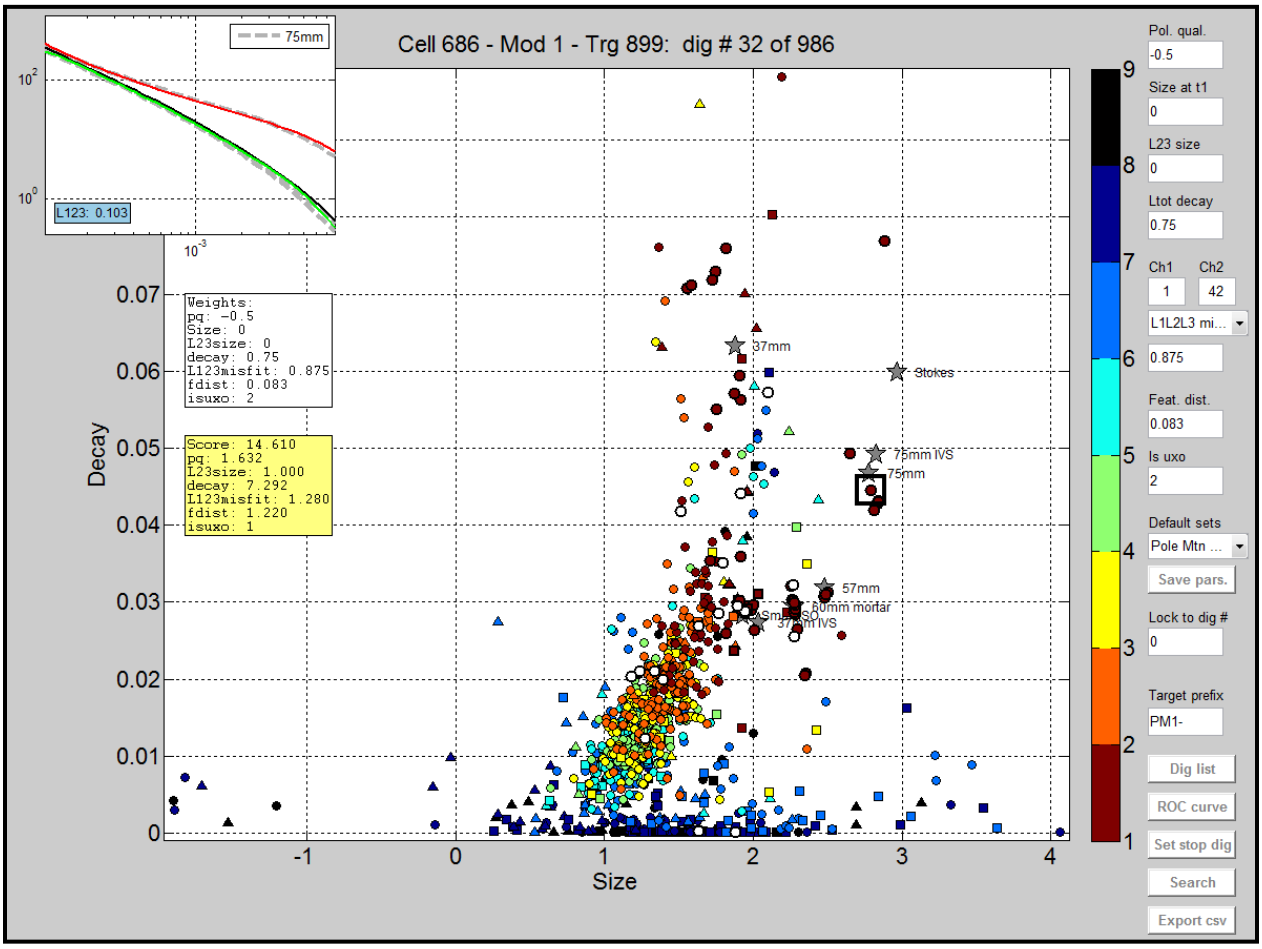


Figure 15. *Dig list tool graphical user interface. Features are plotted in decay-size space with each feature color-coded according to its location in the dig list (red earliest; black latest). Features colored white are training items. Inset at top left shows polarizabilities of the selected anomaly (feature surrounded by a black square). The dig list order is based on the weights shown on the right. These can be specified manually or optimal weights can be determined by a search procedure. The latter approach was used for the Pole Mountain data.*

A stop dig point was determined by visual inspection of the predicted polarizabilities (in relation to the best fitting reference polarizabilities) of each anomaly plotted in dig list order (Figure 16). The stop dig point was conservatively set to the latest anomaly in the dig list with polarizabilities judged to have a realistic possibility of corresponding to a TOI.

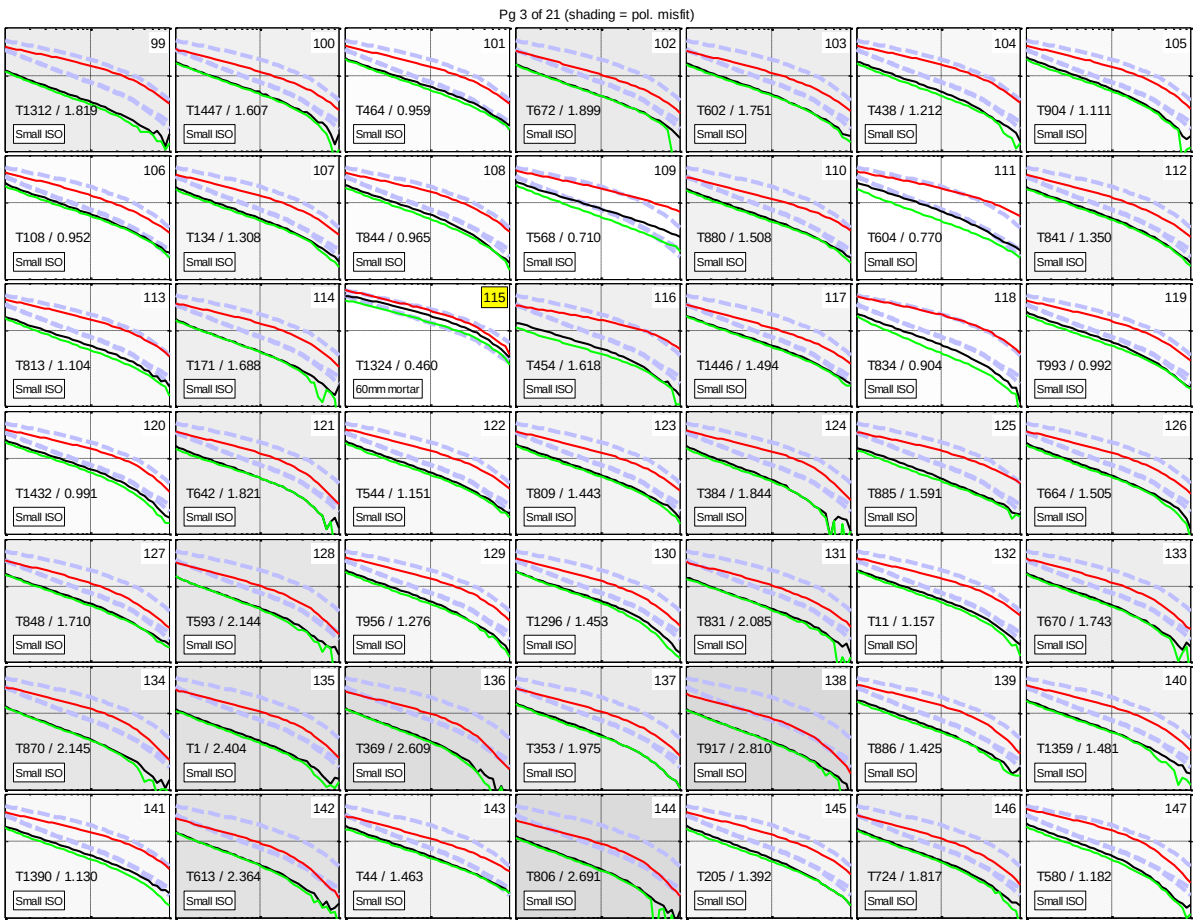


Figure 16. *Example display of polarizabilities plotted in dig list order (for the aggressive dig list). Here, polarizabilities for dig numbers 99 through 147 are shown. Number in top right of each panel is dig number. Numbers in lower left are anomaly number (preceded by "T") and misfit (calculated using all three polarizabilities) to best fitting reference polarizabilities (broken light blue lines). Text in lower left corner is name of closest fitting reference item. Each panel is shaded according to misfit, with larger misfits corresponding to darker shading. Yellow-highlighted dig number (115) is the stop dig point that was chosen for the aggressive dig list.*

Figure 17 shows the partial ROC curve obtained from the program office for the aggressive dig list. This result looked excellent: the partial ground truth showed that the last TOI found was at dig 74, followed by 41 non-TOI digs. The number of digs necessary to achieve a specified confidence level that all TOI have been found can be estimated by fitting a bi-normal distribution to the observed partial ROC curve (Figure 18). At the ~99% confidence level we found no further digs were required for both dig lists. Based on this, and based on the perceived sense that the data were of very high quality capable of providing good constraints on all polarizabilities, no modifications were made to the originally submitted dig lists.

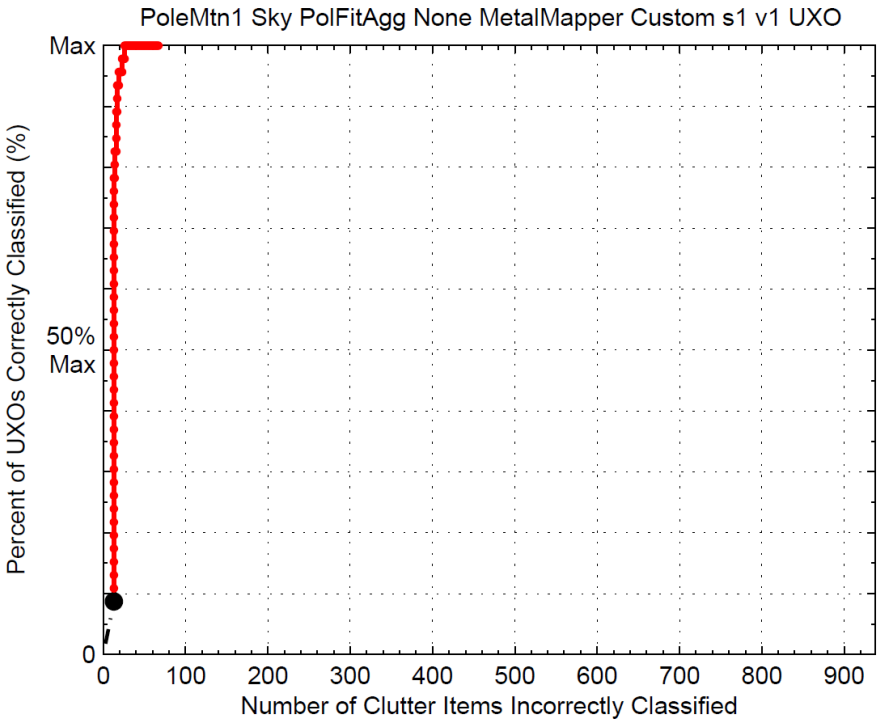


Figure 17. *Partial ROC curve for the year 1 aggressive dig list.*

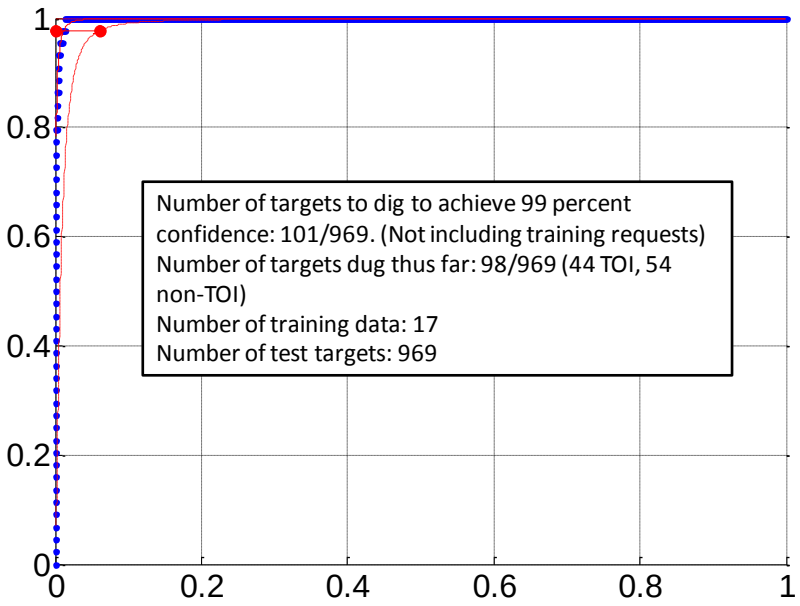


Figure 18. *By fitting a bi-normal distribution to the observed partial ROC curve the number of digs necessary to achieve a specified confidence level (that all TOI have been dug) could be estimated. At the (almost) 99% confidence level, our partial results suggested that no further digging was necessary for the aggressive dig list.*

Year 2 data

The exact same procedure was followed for the year 2 dataset. Training data were requested for an additional five items (Anomalies 1225, 1664, 1780, 2169 and 2033), all of which proved to be non-TOI. An aggressive and not aggressive dig list were submitted. The partial ROC curve for the former is shown in Figure 19. At the 99% confidence level no further digging was required for either list.

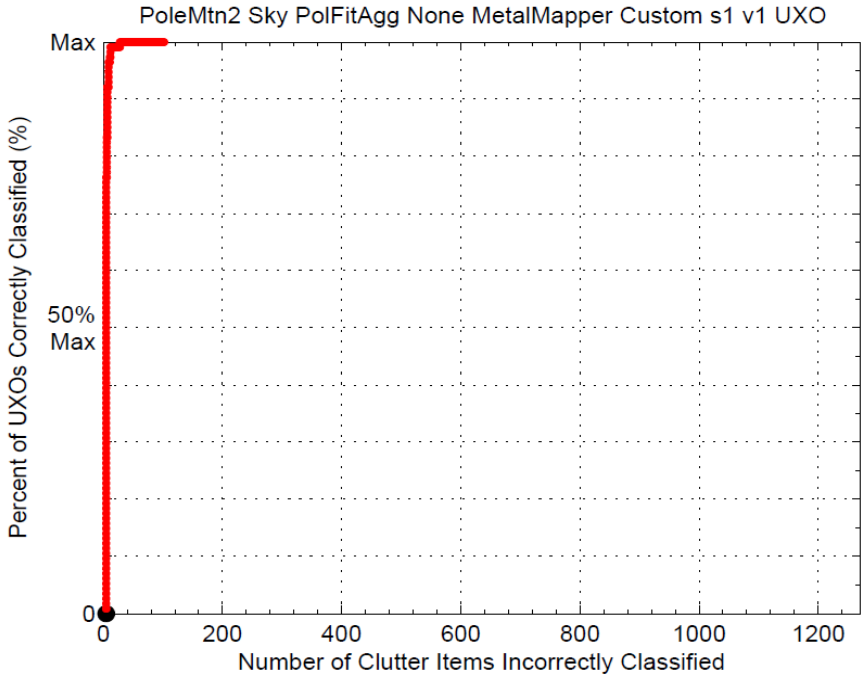


Figure 19. *Partial ROC curve for the year 2 aggressive dig list.*

6.2.3 Retrospective analysis

Figure 20 shows the final ROC curves for the years 1 and 2 datasets. For year 1 all 46 TOI were found after only 26 and 96 non-TOI digs, respectively, for the aggressive and not aggressive dig lists. For year 2 all 114 TOI were found after only 28 and 32 non-TOI digs, respectively.

Feature plots for the years 1 and 2 datasets are shown in Figures 21 and 22. These figures also show the polarizabilities for the last five TOI found with the aggressive dig list. Of particular note is the high quality of these polarizabilities, confirming the overall high quality of the Pole Mountain dataset.

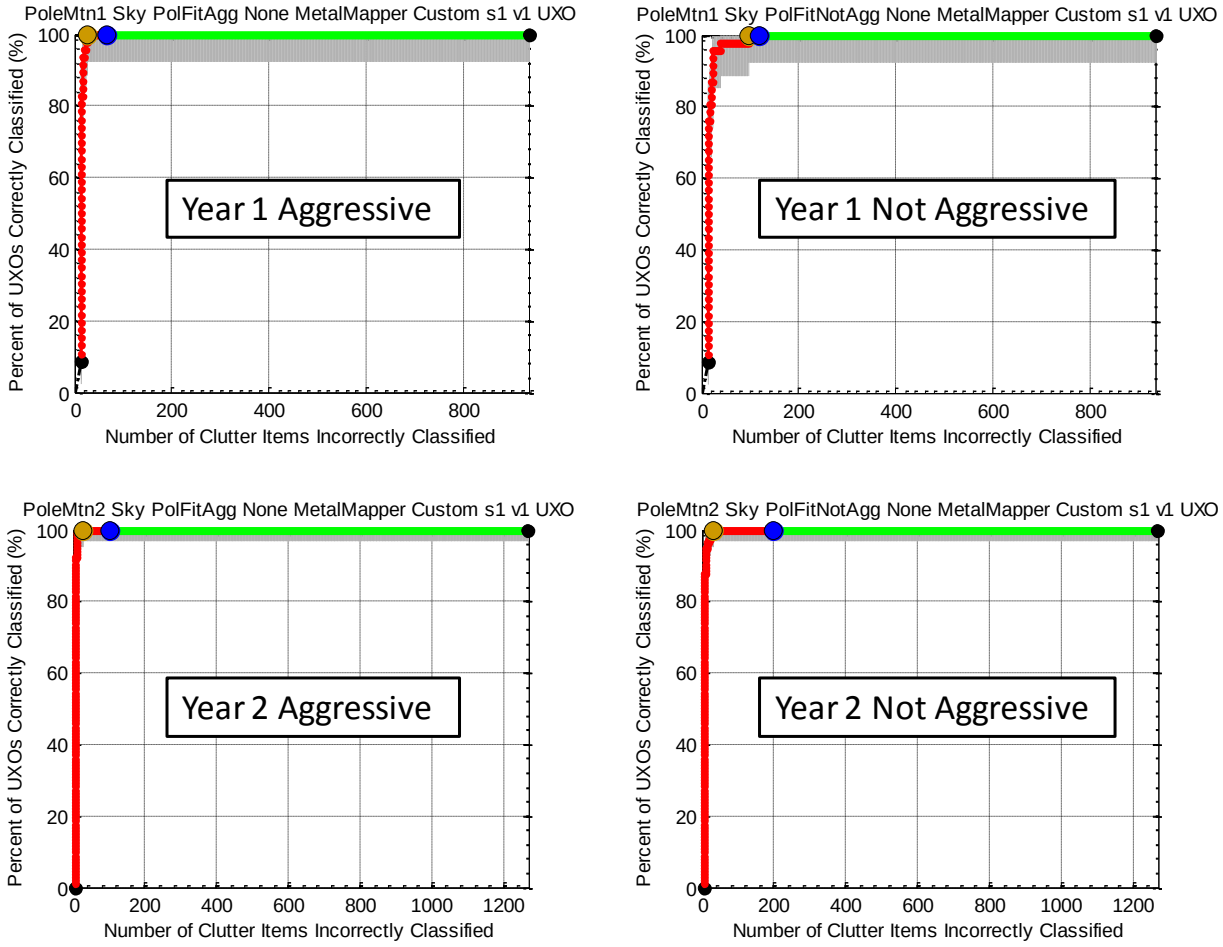


Figure 20. *Final ROC curves for years 1 and 2. All dig lists found all TOI before the stop dig point; there were no difficult TOI.*

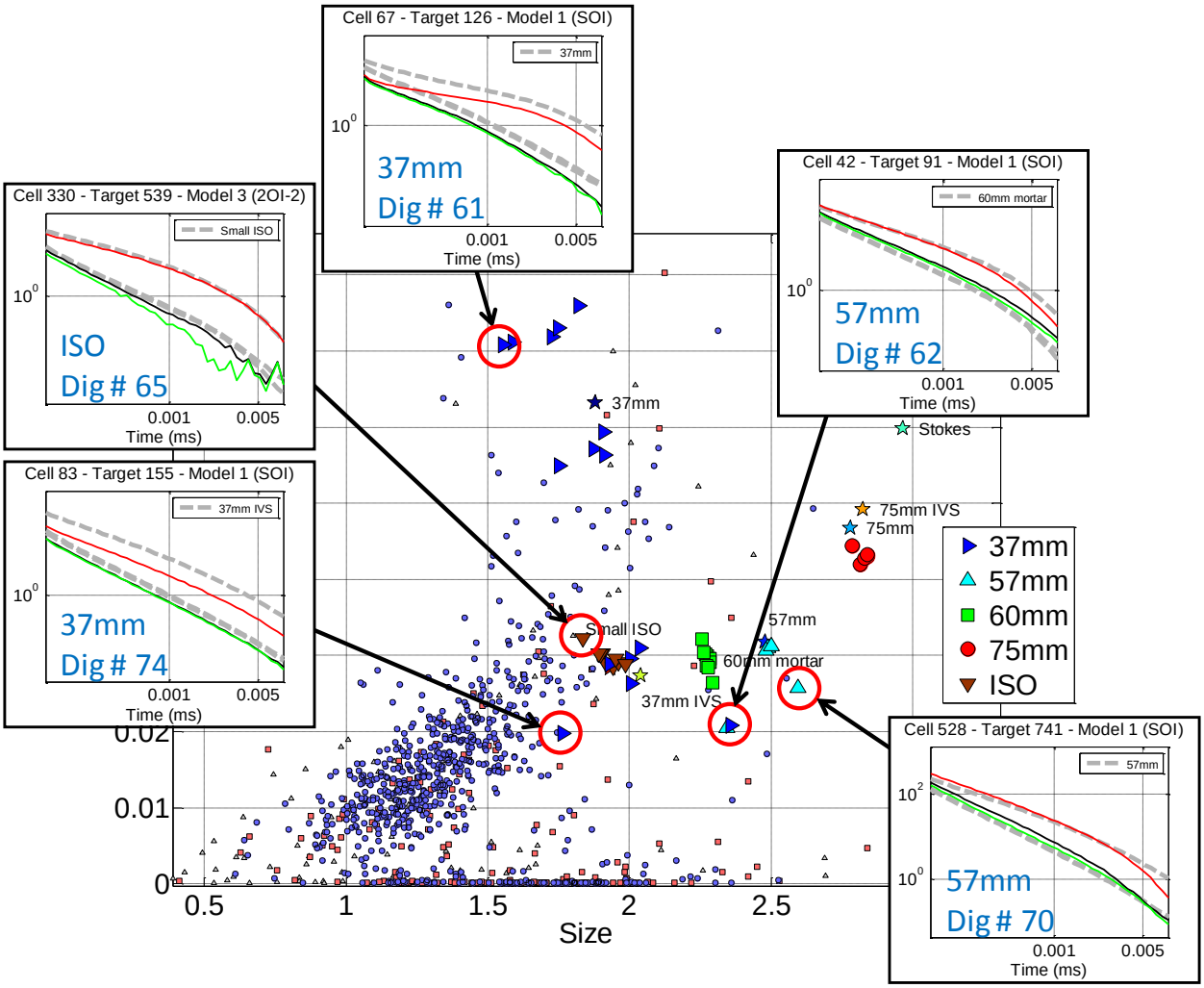


Figure 21. *Decay versus size feature plot for year 1 data showing location of all 46 TOI (large symbols). Small symbols are non-TOI. Labeled stars are reference items. Polarizabilities of the last five TOI found with the year 1 aggressive dig list are shown. Note the high quality of the polarizabilities, even for these, the most "difficult" items.*

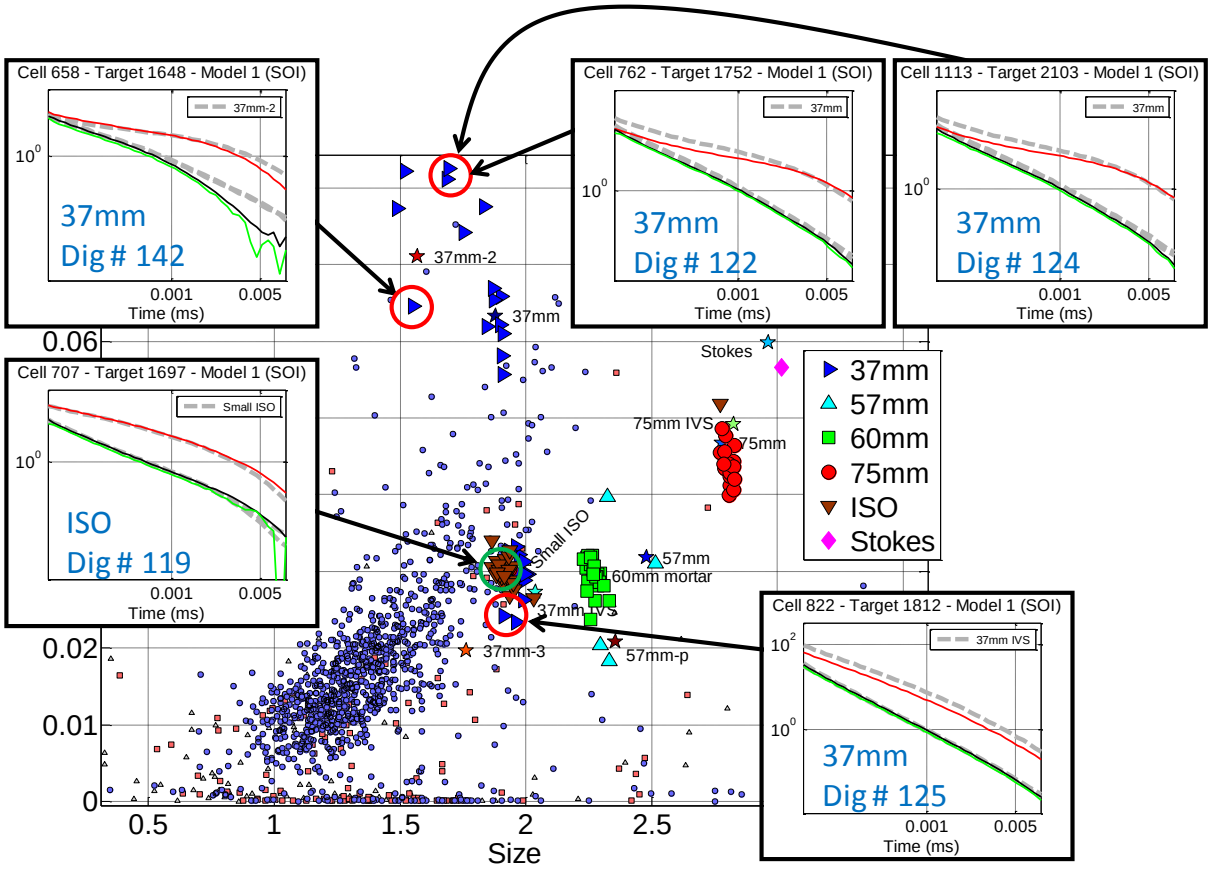


Figure 22. *Decay versus size feature plot for year 2 data showing location of all 114 TOI (large symbols). Small symbols are non-TOI. Labeled stars are reference items. Polarizabilities of the last five TOI found with the year 2 aggressive dig list are shown. Note the high quality of the polarizabilities, even for these, the most "difficult" items.*

Combined years 1 and 2 hypothetical analysis

In this section we look at the performance that would have been obtained if the same analysis procedures used for the year 1 and year 2 datasets had been followed with the combined years 1 and 2 dataset (2370 anomalies). The 22 training items from the separate analyses were treated as training data for the combined dig list. Figure 19 shows the resulting dig list. All 160 TOI are found after 242 digs (80 non-TOI digs). As a testament to the high quality of the data, a dig list based solely on misfit to all three polarizabilities (Figure 23 inset) also performs very well, with all TOI dug after 265 digs (103 non-TOI digs).

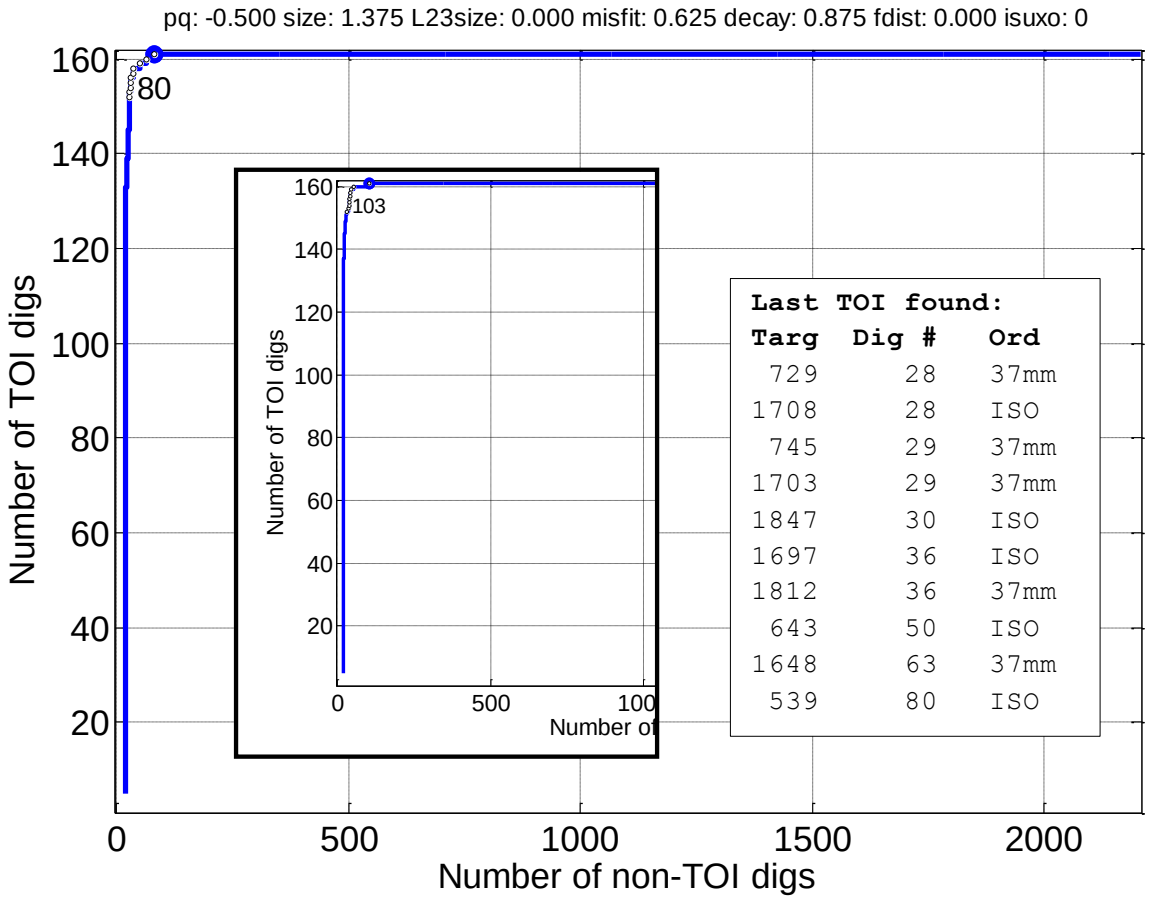


Figure 23. *ROC curve for combined years 1 and 2 data. The dig list was constructed following the same procedure used for the year 1 and year 2 dig lists. All 160 TOI are found after 80 non-TOI digs. Inset shows (part of) a dig list based solely on the misfit of all three polarizabilities to library reference items; all TOI are found after 103 non-TOI digs.*

Figure 24 shows the location in decay versus size feature space of all TOI for the combined dataset. The relatively tight clustering of the TOI in feature space and the separation of most TOI from clutter again attests to the high quality of the data.

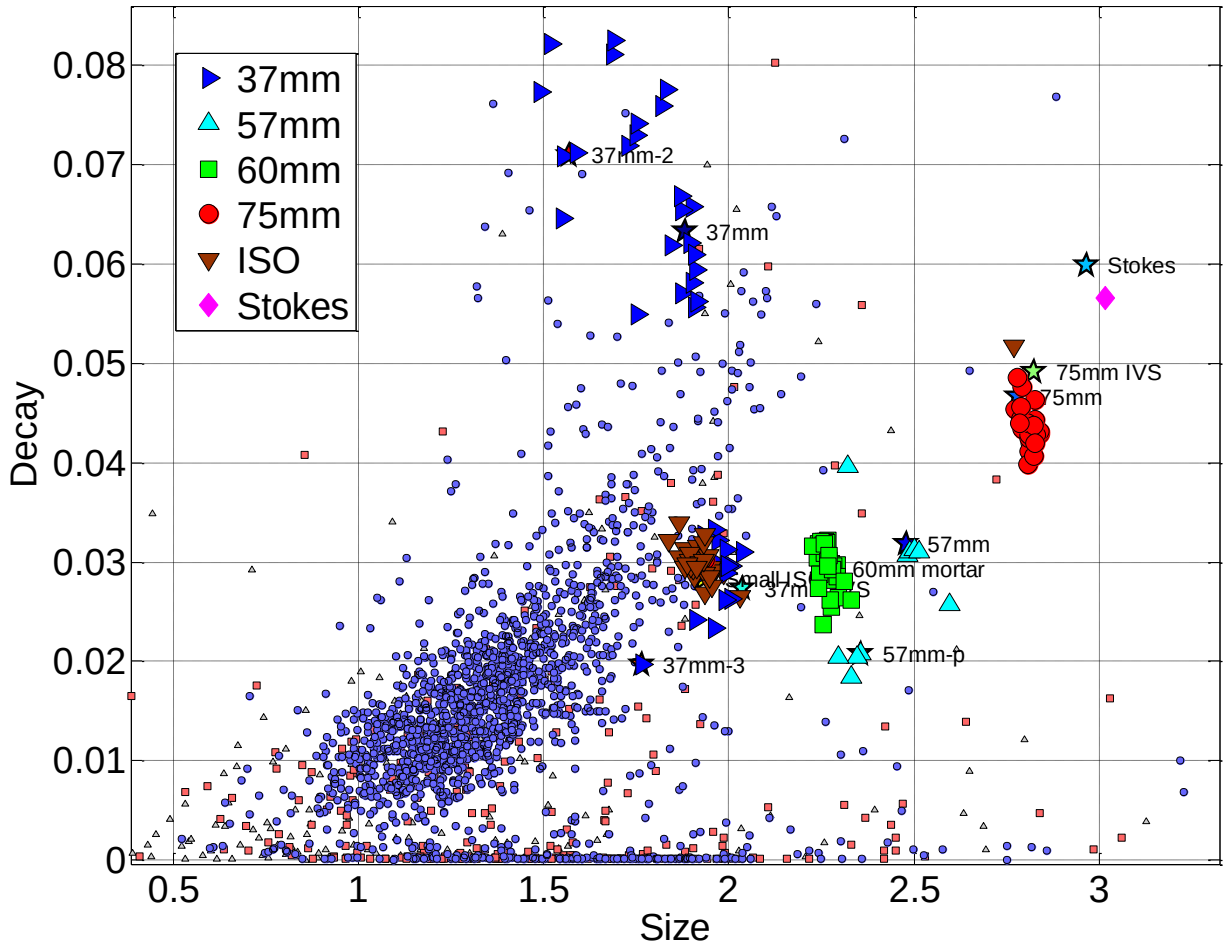


Figure 24. *Decay versus size feature plot for the combined years 1 and 2 datasets. Large symbols show locations of 160 TOI in feature space. Small symbols are non-TOI. Labeled stars are reference items.*

6.3 Support vector machine diglists

For comparison with diglists generated with (some) analyst intervention, we analyzed the Pole Mountain MetalMapper data using support vector machine (SVM) classifiers.

We relied upon the training requests submitted by the analyst for each individual data set, no additional training requests specific to the SVM classifiers were submitted. The training data for the SVM classifiers therefore comprised TOI features from test pit measurements and training requests. When training statistical classifiers, we do not use non-TOI from training requests as these clutter items are typically queried because they are similar to TOI in size-decay, or polarizability, feature space. Including such clutter items in the training data can lead to poor generalization to the test data as the classifier may over fit training non-TOI that are very close to TOI. We therefore prefer to train SVM classifiers using *assumed* non-TOI identified in the test data as follows:

1. We compute a misfit matrix **M** with elements

$$\boldsymbol{M_{jk}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{x_i^j} - \boldsymbol{x_i^k})^{\boldsymbol{2}}. \tag{13}$$

   Here x_i^j denotes the i[th] element of the j[th] training vector, and similarly x_i^k denotes an element of the k[th] test vector. The feature vectors **x** can, in general, be any features derived from the dipole model, here we compute misfits between (log transformed) total polarizabilities

$$\boldsymbol{L_{tot}(t_i)} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{3}} \boldsymbol{L_j(t_i)} \tag{14}$$

   over the full range of channels for each instrument. Alternatively, this operation can also be carried out using size-decay features, though in this case the features must be normalized so that the size feature does not dominate the misfit.
2. We convert each row of **M** to an integer rank vector, producing the rank matrix **R**. That is, the element M_{jk} with minimal misfit over all elements in the j[th] row is assigned the value 1, while the maximal misfit element, for a total of K test vectors, is assigned the value K.
3. The total rank **T** vector is then the column sum of the rank matrix $T_k = \sum_{j=1}^{J} R_{jk}$
4. Finally, we sort the rank vector in descending order to determine a ranking of test feature vectors that are farthest from training TOI in the sense of the misfit function. The first $N_{clutter}$ items in the sorted list are then used as training non-TOI, with $N_{clutter}$ ranging from approximately 100-500 items, depending on the size of the test data set. Note that in this analysis we do not account for the presence of multiple feature vectors for a given target, so that two passed models from the same item could be used in the set of assumed non-TOI.

Figure 25 shows the training TOI and non-TOI feature vectors identified using this approach for the Pole Mountain MetalMapper data. We emphasize that this analysis is carried out using total polarizabilities, for simplicity we display the selected feature vectors in size-decay space.

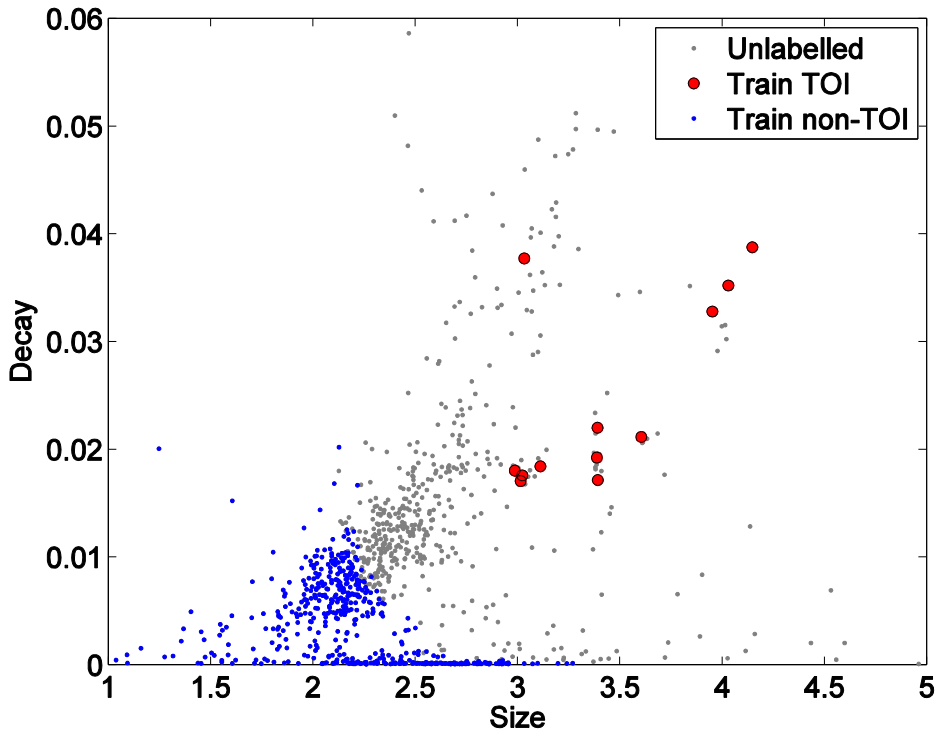


Figure 25. *Size-decay space for Pole Mt. MetalMapper data. Assumed non-TOI are test feature vectors with maximum total polarizability misfit with training TOI.*

Once the initial training data have been defined, we generate diglists from the outputs of nonlinear SVMs with radial basis functions. We employ a "two-stage" classification approach that combines SVM classifiers trained on two feature sets:

1. All polarizabilities. Secondary and tertiary polarizabilities are more susceptible to noise at late times, and so for these parameters we truncate the time range of channels used for classification. For example, with MetalMapper data sets, we use a longer time range of L_1 (channels 1 (0.11 ms) -42 (7.9 ms)), whereas for L_2 and L_3 we restrict classification to channels 1 to 20 (0.78 ms). Selection of time channels can be automated using feature selection algorithms; however, when dealing with limited TOI training vectors we prefer to use analyst judgment when setting these parameters.
2. Total polarizabilities. The range of channels here is typically the same as that used for L_1, since the total polarizability tends to be dominated by the primary polarizability.

Figure 26 shows a representation of the SVM decision function for a two-stage classifier trained on MetalMapper data. Two distinct bands of likely TOI regions are apparent in polarizability feature space, corresponding to large (75 mm), and small (37 mm, ISO) ordnance, respectively.

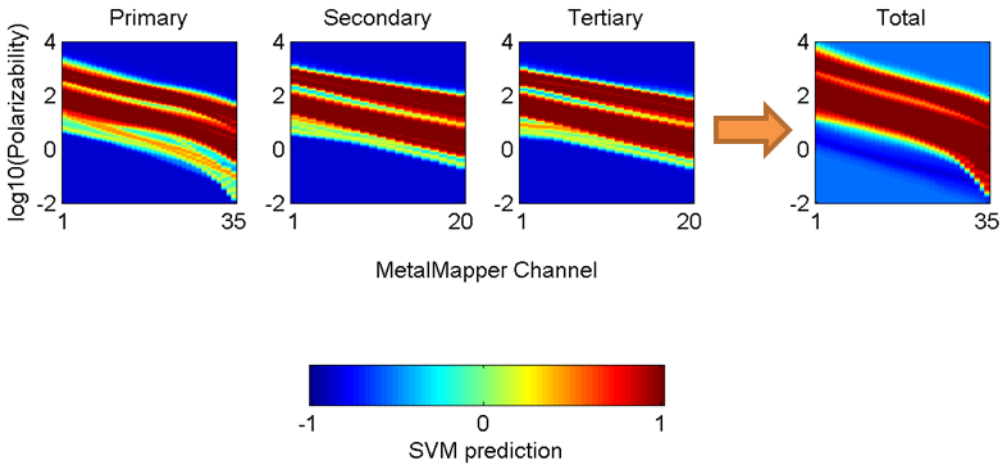


Figure 26. *Decision surface for two-stage SVM classifier applied to MetalMapper data. An SVM prediction of 1 indicates a high likelihood of TOI.*

For both feature sets, the polarizabilities are log-transformed prior to training and prediction stages. We use a width of σ=1 for the Gaussian kernels; this choice is based on past applications of this algorithm to polarizability features. Cross validation techniques can also be used to select the kernel width.

Our two-stage approach is motivated by the observation that the majority of targets of interest interrogated with next generation sensors produce well-constrained polarizabilities that are an excellent match to training TOI. In the initial stages of digging we therefore wish to exploit all available information in the recovered model and should use all polarizabilities in discrimination. However, a small proportion of TOI can still produce poorly constrained transverse polarizabilities, particularly if the sensor is not properly centered relative to the target. In these cases, we find that the total polarizability can still match training vectors. However, in this second stage the false alarm rate will inevitably begin to increase as clutter can often match total polarizabilities for smaller TOI (e.g. 37mm or ISOs). We therefore expect that the ROC for two-stage SVM classification will rise sharply and then "turn over" slightly in the second stage. This strategy was successfully employed for classification of Camp Butner MetalMapper and TEMTADS data. Retrospective analysis on Camp Butner data sets showed that the two-stage approach reduced the significantly final false alarm rate relative to a single SVM that relies on all polarizabilities throughout digging.

When should we make the switch between the first (all polarizabilities) and second (total polarizabilities) stages of digging? A plot of the decision statistic f_{SVM} sorted in descending order has an inflection point where the first stage classifier transitions from clear matches to known TOI to poorer matches (Figure 27). In Figure 27 this corresponds to a value of f_{SVM}≈-0.5, indicating we are digging halfway between the decision boundary (f_{SVM}=0) and the support vectors for the non-TOI class (f_{SVM}=-1). We select this point to transition to the second stage classifier. A similar inflection point is identified in the second stage as our initial stop dig point.

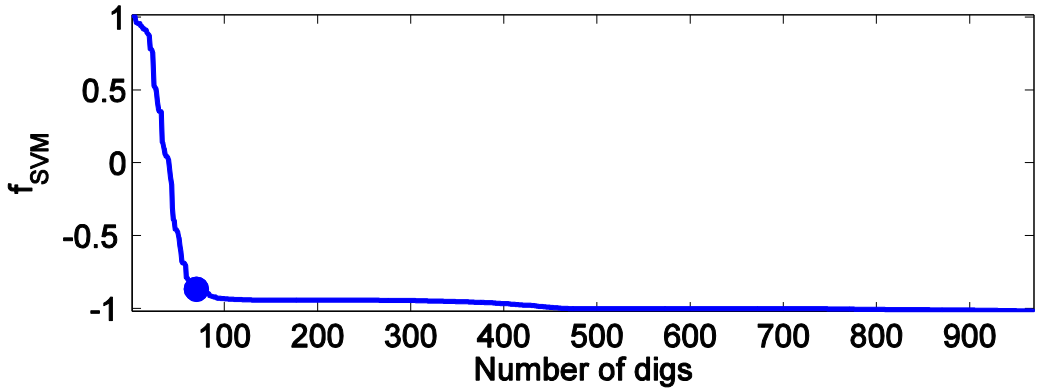


Figure 27. *SVM decision statistic f_{SVM} for stage 1 (all polarizabilities) SVM classifier applied to Pole Mt. Year 1 MetalMapper test data. Marker indicates point in dig list at which we switch to stage 2 (total polarizability) classifier.*

Once groundtruth is received up to our initial stop dig point, we update the SVM diglists by either retraining with newly encountered TOI, or, if necessary, extending the stop dig point to achieve a specified confidence that no more TOI remain in the ground. We describe these procedures in the following sections.

6.3.1 Selecting a stop dig point

If no novel TOI are encountered up to the current stop dig point, then we have some confidence that the training data have adequately characterized the test data and no TOI will be left in the ground with the current classification strategy. To formalize this determination, we use a binormal model of the ROC to arrive at a final stop dig point. This model assumes that the observed ROC can be represented as a sample from two normally-distributed score distributions. The resulting ROC curve is a function of two parameters (Metz et al, 1998)

$$\boldsymbol{a} = \frac{|\boldsymbol{\mu_1} - \boldsymbol{\mu_2}|}{\boldsymbol{\sigma_1}} \qquad (15)$$

$$b = \frac{\sigma_1}{\sigma_2}.$$

Maximum likelihood estimation of these parameters yields the predicted ROC and the model covariance, which can then be used to determine confidence intervals on the ROC. In practice, the generating score distributions of TOI and non-TOI are rarely normally distributed, and so estimating means and variances directly from the empirical score distributions will yield a poor fit to the observed ROC. A better strategy is to express the predicted ROC as a function of the parameters *a* and *b* and to fit the observed ROC directly. This approach can often yield an excellent fit, even if the underlying score distributions are not normally distributed (Hanley, 1988). This is because the observed ROC is invariant under arbitrary monotonic transformations of the decision statistic (i.e. transformations that preserve the ordering of the diglist).

We use this model to test whether all ordnance have been found at a selected dig point as follows:

1. Fit a binormal model to known ground truth (i.e. up to selected stop dig point), assuming as our null hypothesis that the number of detected targets of interest N_{TOI} at the site is the number of TOI found thus far.

2. Determine a confidence interval on the expected false alarm rate by finding the point at which the estimated binormal confidence interval has the value

$$\boldsymbol{P = 1 - 1/N_{TOI}.} \tag{16}$$

For the data in Figure 28 this interval is shown as a solid horizontal line.

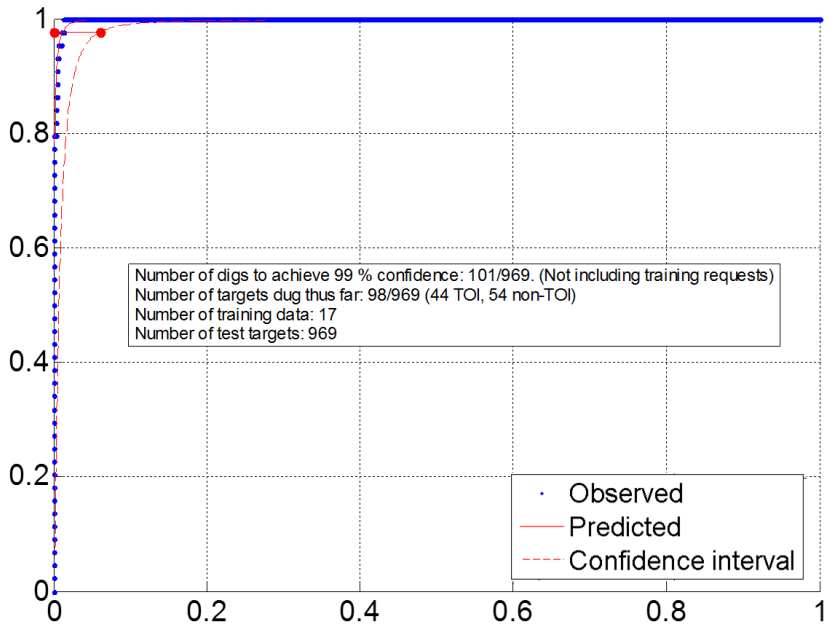


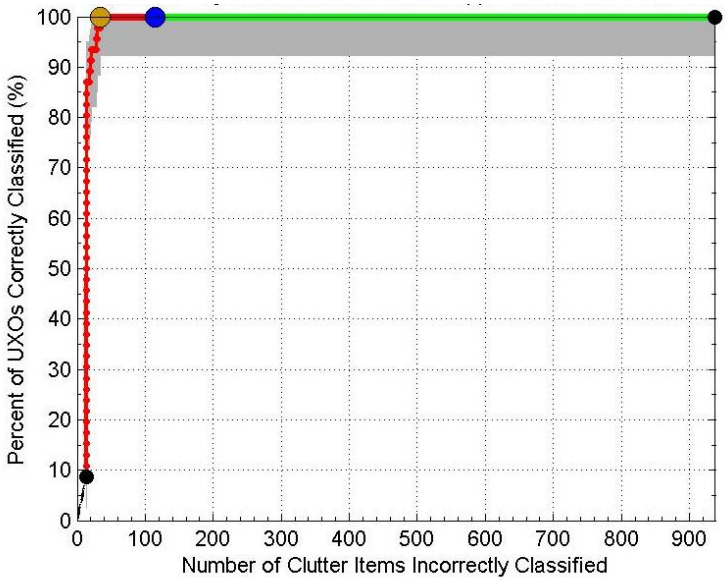
Figure 28. *Predicted binormal ROC and 99% confidence interval for Pole Mt. Year 1 MetalMapper polarizability match dig list. An additional 3 digs (from 98 to 101) are required to test the null hypothesis.*

We then test our null hypothesis at the specified confidence by digging out to the maximum extent of the confidence interval on the FAR. If no further TOI are encountered then we retain the null hypothesis and finish digging. If new TOI classes *are* encountered then we retrain our classifier. Retraining was not required for the Pole Mountain MetalMapper data.

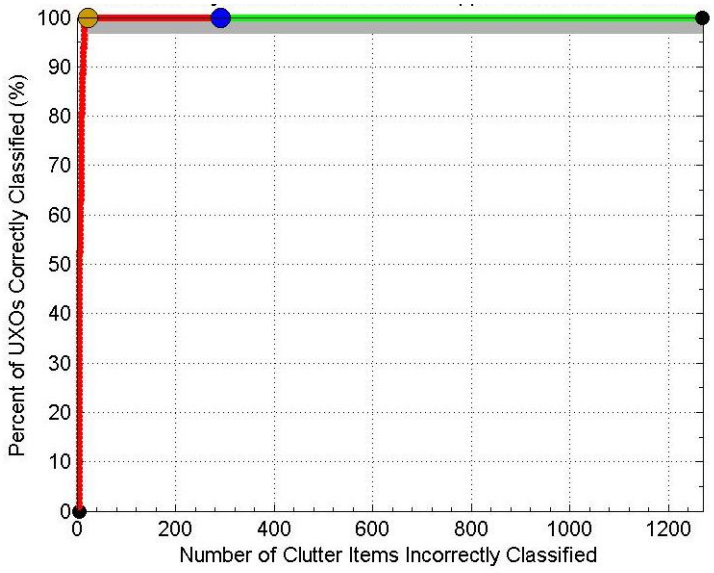
6.3.2 Retrospective analysis of SVM classification at Pole Mountain

Excellent classification performance was obtained with the SVM for both years of the Pole Mountain demonstration (Figure 29). The keys to successful discrimination are reliable polarizability estimates and sufficient training data; both requirements were met at this site.

The only criticism of this result is that our stop dig point, initially selected from the inflection point of the SVM decision statistic, extended far past the last TOI for the Year 2 data. This problem can be easily addressed by selecting an initial, aggressive stop dig point by visual inspection of the dig list, followed by a bi-normal fit of the partial ROC to assign a numerical confidence to the final stop dig point, as described above.



(a) Year 1 data



(b) Year 2 data

Figure 29. *ROC curves for SVM classification of Pole Mountain Year 1 (top) and Year 2 (bottom).*

6.4 Technology Transfer with Shaw Environmental

The objective of the project was to transition mature data analysis and evaluation techniques and specialized knowledge base regarding dipole-based advanced discrimination from Sky Research to Shaw's production team through a one week training session with Sky's technology developers in Vancouver, Canada.

The geophysical datasets used for the training consisted of dynamic EM61-MK2 measurements with integrated GPS and static MetalMapper records for 2,370 anomalies at the Pole Mountain Demonstration Site located in Wyoming.

Sky provided background information and general training on their UXO Lab software routines and data inversion and classification techniques (Section 2). Shaw used the UXO Lab software routines and Sky information presented during training to analyze 2,370 anomalies and output a diglist of suspected target off interest (TOI).

6.4.1 Initial Data Screening and QC

MetalMapper records for 2,370 locations were imported into UXOLab and an inversion was performed to generate information on the x-y-z location and polarizabilities for each record. Both single and multi-target models were fitted during the inversion process and these data were subsequently analyzed using the *QC Tool Flex* routine in UXOLab. The initial evaluation allowed the analyst to review MetalMapper polarizability profiles for each record in conjunction with the amplitude and spatial attributes of each EM61-MK2 anomaly, inversion model fit statistics (signal to noise ratios, general model uncertainty, uncertainty of recovered polarizability, color-coded images of the polarizabilities for all 9 transmitter-receiver combinations) and predicted depth for each model fit. Of particular importance were the polarizability curves from the TOI from the library reference items that were superimposed on polarizability profile of each MetalMapper record, which allowed a direct comparison of the similarity of the polarizabilities. A decay-size feature plot was also useful for evaluating the current model's attributes compared to the entire dataset and library of reference items.

Each model and inversion result was passed or failed by the data analyst. Additionally, notes regarding whether the anomaly was a suspected TOI for the project based on the library of reference items or an elongated, UXO-like object were made in order to facilitate final ranking during diglist development.

6.4.2 Data Analysis and Selection of Training Data

The models passed by the analyst from the inversion process were displayed on a feature plot with the decay on the y-axis and relative size on the x-axis along with the library reference items (37 mm (2 types), 57 mm, Stokes Mortar, 75mm (2 types), and small ISO). Anomalies that clustered close to the library reference items were evaluated further in terms of the similarity of their polarizabilities to those of the known TOI. During the analysis and evaluation process, other clusters (or populations) of anomalies were identified, some of which exhibited UXO-like polarizabilities, signal amplitude, and decay properties upon examination. Examples are 1) one large cluster with "sub-clusters" of smaller relative sizes than library reference items, variable decay rates, and UXO-like polarizabilities, 2) relative size larger than a small ISO and smaller than a 57mm with UXO-like polarizabilities and decay properties,

and 3) non-clustered but having interesting combinations of polarizability, decay, or relative size characteristics. Shaw utilized these populations to select training data using Sky's *QC Training* module.

A total of 22 anomalies were selected for training (Table 4). The first set of data requested were limited to the Year 1 data since the QC of the Year 2 intrusive results was in progress. Five anomalies were selected and 2 resulted in the identification of different types of 37mm projectiles. The second round of training data (11 anomalies) resulted in 1- 57mm projectile, 2- 60mm mortars and 1- 75mm projectile. The final training request for 6 anomalies resulted in 1- 60mm mortar. It is of particular interest that 60mm mortars were identified using the training process. Prior to this development, 60mm mortars were not considered a TOI for the project based on the library of reference items for this demonstration. The non-TOI items selected as part of the training process resulted in elongated, UXO-like items that were either MD or CD.

Table 4: *Training Data Request and Results: First Training Request*

Training Request Number	Anomaly ID	Depth (cm, measured)	Dip Angle	Azimuth	Identification	Length	Dig Type
1	PM-164	5	0	20	Frag (light)	5	MD
	PM-802	4	0	320	Frag (light)	10.5	MD
	PM-1340	2	0	330	Other	32	CD
	PM-126	21	0	350	37mm projectile	9	TOI
	PM-155	18	0	90	37mm projectile	9	TOI
2	PM-443	4	-80	0	Frag (light)	6	MD
	PM-1231	4	0	70	Frag (heavy)	12	MD
	PM-1646	4	0	270	Frag (medium)	13.5	MD
	PM-1709	30	-85	120	60mm mortar	13	TOI
	PM-2112	20	0	0	75mm projectile	23	TOI
	PM-1649	25	-85	160	60mm mortar	13	TOI
	PM-2098	35	0	0	Frag (light)	21.5	MD
	PM-2098	35	0	0	Frag (light)	7.5	MD
	PM-2098	35	0	0	Frag (light)	6	MD
	PM-2098	35	0	0	Frag (light)	4	MD
	PM-2098	35	0	0	Frag (light)	3.5	MD
	PM-1354	8.5	0	0	Frag (medium)	5	MD
	PM-1104	3	15	270	Frag (medium)	13	MD
	PM-988	20	0	180	57mm projectile	12	TOI
	PM-144	4	0	0	Frag (light)	15	MD
3	PM-194	7	-90	0	Fuse/Fuse Components	6.5	MD
	PM-851	16	60	0	Frag (medium)	14.5	MD
	PM-851	14	0	0	Frag (light)	13.5	MD
	PM-851	16	60	30	Frag (medium)	18.5	MD
	PM-851	22	0	0	Frag (light)	11	MD
	PM-851	2	0	0	Frag (light)	4	MD
	PM-895	2	10	90	Frag (medium)	12	MD
	PM-1665	8	0	90	Frag (medium)	11.5	MD
	PM-277	30	-80	90	60mm mortar	13	TOI
	PM-2141	4	0	90	Frag (light)	10	MD

During this phase of the project the 60mm mortar polarizability information was added to the UXOLab TOI library by Sky for use in the automated development of the final diglist.

6.4.3 Initial Diglist Development

Initial diglist development was accomplished using the Sky *Digzilla* tool, which is designed to automate the prioritization of the diglist. Using this tool, the analyst is able to interactively prioritize anomalies using various parameters such as polarizability misfit and quality, decay, relative size, analyst notes (e.g., "UXO-like") and other related attributes. Shaw used the minimum polarizability misfit compared to the library of reference items and analyst notes (e.g., , "UXO-like") as the primary attributes to refine the initial diglist.

The training data results were also automatically ranked using *Digzilla*. Their auto-generated rankings were scattered within the highest 1750 priorities and 12 would have fallen above the "stop dig" point that was automatically selected by *Digzilla*. This result was somewhat expected since the purposes of the training data were to 1) discount anomalies that did not cluster and / or have properties similar to the library of reference items falling within specific clusters in the feature plot and 2) investigate anomaly populations that had properties that were UXO-like. Shaw used the results of the initial *Digzilla* output and subsequently employed an iterative process to refine select *Digzilla* parameters with the assistance of Sky personnel.

The prioritized diglist output from *Digzilla* was reported in MatLab and transferred into an Excel spreadsheet for final prioritization.

6.4.4 "Stop Dig" Point Selection and Final Diglist Prioritization

This project required that anomalies on the diglist be categorized as "likely UXO", "can't decide", "likely clutter" or "can't extract reliable parameters". Since the overall data quality was considered to be very good with regards to the inversion results for the field data and library of reference items, Shaw classified all of the anomalies on the diglist as either "likely UXO" or "likely clutter". Each anomaly was then marked as "dig" or "no dig".

Based on the results of *Digzilla* and analyst notes, approximately 180-210 anomalies were of interest during generation of the original diglist. The *QC Training* tool was employed to automate the selection of the "stop dig" point by defining the confidence level, the anomalies selected as "dig" and the total number of anomalies. After the results of the first iteration of the diglist were reviewed a 99% confidence level was set and the *QC Training* tool suggested ~ 60 additional anomalies for investigation would be appropriate. Shaw used the number of anomalies specified by the *QC Training* tool as a guide and manually reviewed the anomalies surrounding the recommended stop dig point with the *QC Flex tool.* This procedure was performed to ensure the classification was optimized for each anomaly based primarily on the polarizabilities, decay, depth, and relative size. During the manual review, 8 additional anomalies were classified as "UXO-like" to ensure that all anomalies thought to be TOI were investigated.

A portion of the ~ 60 anomalies reviewed were characterized by inversion results that had one or more parameter estimates thought to be unrealistic for the specified model (e.g., large depth estimate, predicted location at edge of search window, or relatively noisy polarizability curves). Because the data

analyst was uncertain and wanted to err on the side of conservatism, some of these anomalies were reclassified as “likely clutter”, “dig”. Based on the results sent to SERDP, no TOIs were present in the additional selections and no new information was gained and it was decided to stop “digging”.

Figure 30 summarizes the overall results of the dig selection process.

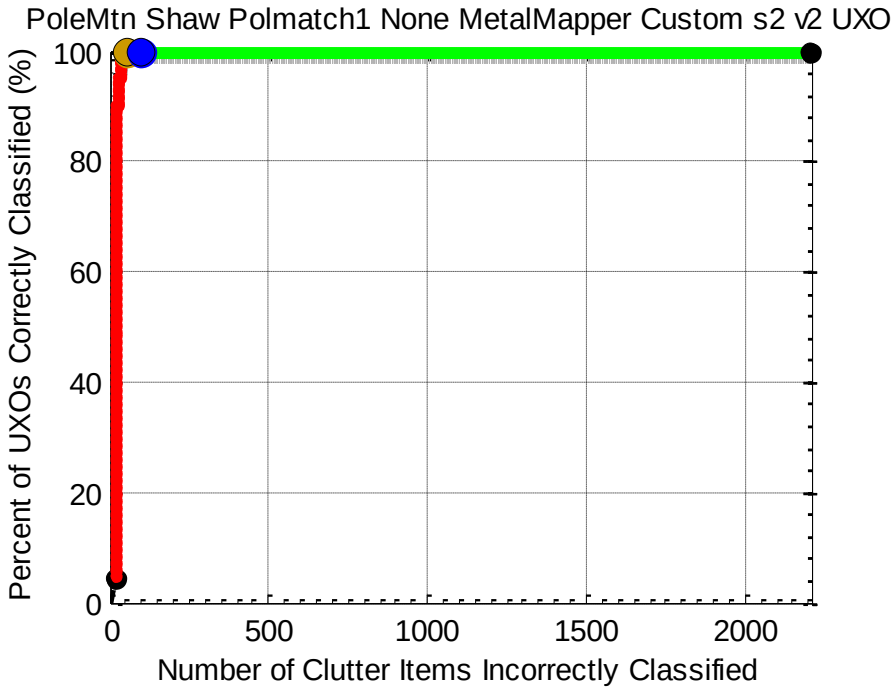


Figure 30. *ROC curve for the dig selection process*

6.4.5 Summary

Shaw geophysicists were trained in the use of Sky’s discrimination and classification methods using several UXOLab modules (*QCFlex, QCTraining*, and diglist prioritization using *Digzilla*).

Of the 22 targets selected as training data 7 were TOI, including a 60mm mortar that was not present in the original library of reference items. 15 were UXO-like and generally smaller than the smallest item in the library of reference items.

Initially, 200 of the 2,370 anomalies were selected for investigation. The information attained from the training data and *QC Training* module resulted in an additional 36 selections for a total of 236 anomalies to investigate (10% of the total number of anomalies).

All 154 UXO (excluding training data) and ISOs were identified with the 236 anomaly selections. The last TOI was ranked as dig 192 on the final diglist.

Very few non-TOI are present amongst the high priority anomalies. The frequency of non-TOI increases with decreasing rank on the diglist.

Appendix 1: Summary of Performance for all methods applied at Pole Mountain

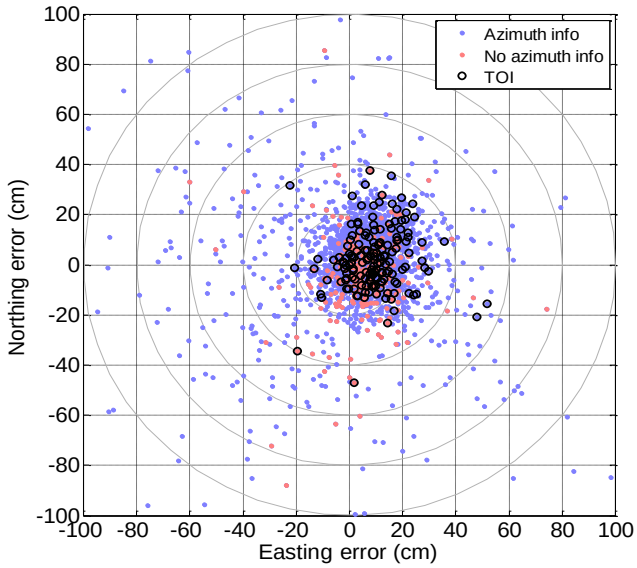
Table 5. *Summary Of Diglists submitted by Sky Research and Shaw/Sky*

Data Set	Instrument	Method	Features	Number of training requests	No. digs	FAR (No. scrap dug at final TOI)	FAR (%)	No. scrap dug at operating point	Scrap Dug at operating point (%)
Year 1 Anomalies = 984 No. Scrap = 938 No. UXO = 46	MetalMapper	Aggressive Library Match on Polarizabilities	Polarizabilities	17	113	26	2.77%	67	7.14%
	MetalMapper	Non-aggressive Library Match on Polarizabilities	Polarizabilities	17	165	96	10.23%	119	12.69%
	MetalMapper	SVM on Polarizabilities	Polarizabilities	17	160	34	3.62%	114	12.15%
	EM61	PNN on decay and data amplitude	Polarizability decay and Data amplitude	0	650	600	63.97%	604	64.39%
Year 2 Anomalies = 1384 No. Scrap = 1270 No. UXO = 114	MetalMapper	Aggressive Library Match on Polarizabilities	Polarizabilities	5	216	28	2.20%	102	8.03%
	MetalMapper	Non-aggressive Library Match on Polarizabilities	Polarizabilities	5	311	32	2.52%	197	15.51%
	MetalMapper	SVM on Polarizabilities	Polarizabilities	5	405	22	1.73%	291	22.91%
	EM61	PNN on decay and data amplitude	Polarizability decay and Data amplitude	1	1059	863	67.95%	945	74.41%
Entire Data Set	MetalMapper	Aggressive Library Match - Technology Transfer w/ SHAW	Polarizabilities	22	258	53	2.40%	97	4.40%

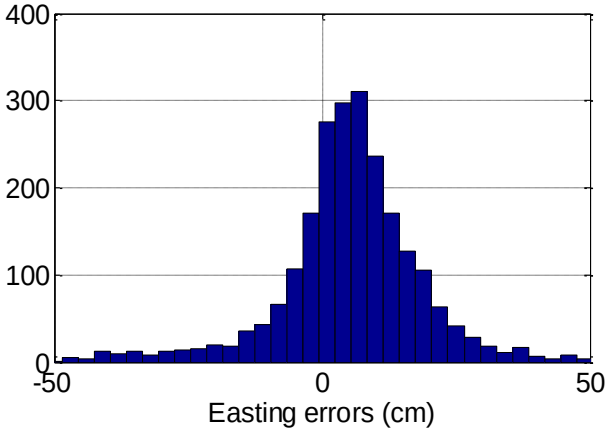
Appendix 2: Target Location Error Analysis for MetalMapper dataset

The recovered locations of targets were compared to the locations reported in the ground truth files provided by the ESTCP program office. For this analysis, we considered the MetalMapper dataset only. Figure A-1 plots the location estimate errors in both Northing and Easting.

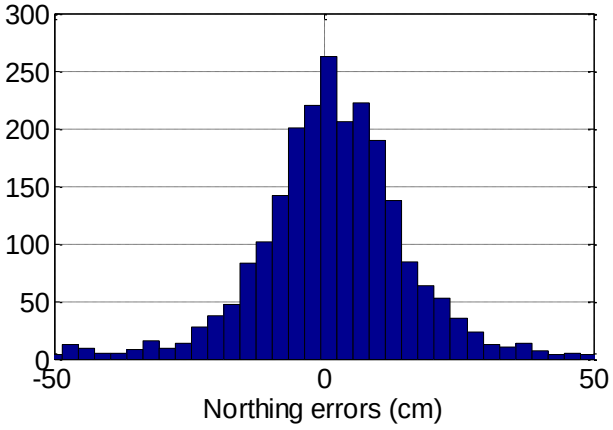
In Figure A-1(a), a scatter plot shows the there is a bias of 5.5 cm in Easting error and 1.4 cm in Northing error. Different colored symbols are used to indicate which anomalies were acquired without azimuth information. In Figures A-1(b) and (c) we plot histograms of the errors. Within an error of +/- 50 cm, the error distribution is approximately Gaussian with standard deviations in Easting and Northing of 13.2 cm and 13.9 cm, respectively.



(a) Location errors. Median error in Easting = 5.5 cm. Median error in Northing = 1.4 cm



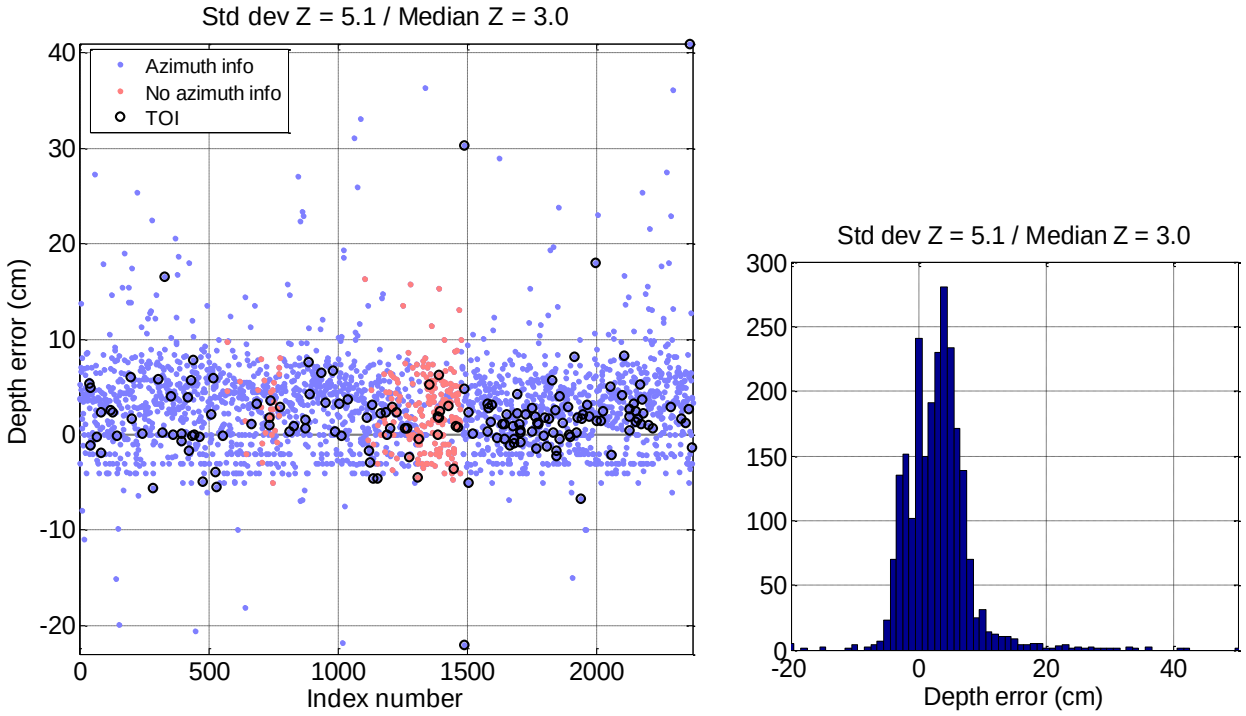
(b) Distribution of errors in Easting. Standard Deviation σ=13.2 cm



(c) Distribution of errors in Northing. Standard Deviation σ=13.9 cm

Figure A-1. *Analysis of location errors*

Figure A-2 plots the distribution of depth estimate errors. The median of the errors is 3.0 cm, and the standard deviation of the distribution is 5.1 cm.



(a) Depth estimate error for all targets

(b) Histogram of depth errors

Figure A-2. *Analysis of depth estimate errors for the MetalMapper Inversions*

Appendix 3: Points of Contact

POINT OF CONTACT	ORGANIZATION Name Address	Phone E-mail	Role in Project
Leonard Pasion	Sky Research Inc, 112A/2386 East Mall Vancouver, BC, V6T-1Z3	541 552 5186 leonard.pasion@skyresearch.com	Principal Investigator (PI)
Kevin Kingdon	Sky Research Inc, 112A/2386 East Mall Vancouver, BC, V6T-1Z3	541 552 5187 kevin.kingdon@skyresearch.com	Project management and personnel coordination
Erik Russell	Sky Research Inc, 3 Schoolhouse Lane Etna, NH 03750	541 552 5197 erik.russell@skyresearch.com	Cost tracking

References

T. Bell, B. Barrow, J. Miller, and D. Keiswetter. Time and frequency domain electromagnetic induction signatures of unexploded ordnance. Subsurface Sensing Technologies and Applications,2:153-175, 2001.

S. D. Billings, L. R. Pasion, L. Beran, N. Lhomme, L. Song, D. W. Oldenburg, K. Kingdon, D. Sinex, and J. Jacobson. Unexploded ordnance discrimination using magnetic and electromagnetic sensors: Case study from a former military site. Geophysics, 75:B103-B114, 2010.

J.A. Hanley, 1988. The robustness of the "binormal" assumptions used in fitting ROC curves. Medical Decision Making, 8: 197-203.

X. Liao and L. Carin. Migratory logistic regression for learning concept drift between two data sets with application to UXO sensing. IEEE Trans. Geosci. Remote Sensing, 47:1454 -1466, 2009.

C. E. Metz, B. A. Herman, and J.H. Shen. 1998. Maximum likelihood estimation of receiver operating characteristic curves from continuously-distributed data. Statistics in Medicine, 17: 1033-153.

L. R. Pasion, S. D. Billings, D. W. Oldenburg, and S.Walker. Application of a library-based method to time domain electromagnetic data for the identification of unexploded ordnance. Journal of Applied Geophysics, 61:279-291, 2007.

F. Shubitidze, K. O'Neill, S. A. Haider, K. Sun, and K. D. Paulsen. Application of the Method of Auxiliary Sources to the Wide-Band Electromagnetic Induction Problem. IEEE Trans. Geosci. Remote Sensing, 40:928-942, 2002.

S. L. Tantum, Y. Li, and L. M. Collins. Bayesian mitigation of sensor position errors to improve unexploded ordnance detection. IEEE Geosci. Remote Sensing Letters, 5:103{107, 2008.

G. F. West and J. C. Macnae. Electromagnetic methods in applied geophysics, chapter Physics of the electromagnetic exploration method, pages 5-45. SEG, 1991.

Y. Zhang, X. Liao, and L. Carin. Detection of buried targets via active selection of labeled data: Application to sensing subsurface UXO. IEEE Trans. Geosci. Remote Sensing, 42:2535-2543, 2004.